HSE
Health & Safety Executive

Frequencies of diseases presenting to General Practitioners according to patients' occupation

Prepared by the
Institute of Occupational Medicine
for the Health and Safety Executive

CONTRACT RESEARCH REPORT
340/2001

HSE
Health & Safety Executive

Frequencies of diseases presenting to General Practitioners according to patients' occupation

C A Soutar
Institute of Occupational Medicine
8 Roxburgh Place
Edinburgh
EH8 9SU
Scotland
United Kingdom

General practitioner morbidity recording schemes offer an opportunity to describe the frequencies of a wide range of conditions. The aim of the study was to examine the feasibility of assembling information nationally, on the frequencies and distribution of ill-health presenting to general practitioners, in relation to occupation.

The study showed, as expected, that none of the Schemes routinely record occupation, but additional procedures to obtain this information are feasible. Together the four main ongoing Schemes can provide a population of sufficient size to detect differences in frequency of ill-health between most occupations. The procedures and data collection are sufficiently comparable to allow the results to be aggregated in some way.

Possible study designs are discussed. The main ethical concern is that there should be no possibility of identification of individual subjects, and methods of satisfying this requirement are suggested.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not necessarily reflect HSE policy.

HSE BOOKS

© *Crown copyright 2001*
Applications for reproduction should be made in writing to:
Copyright Unit, Her Majesty's Stationery Office,
St Clements House, 2-16 Colegate, Norwich NR3 1BQ

First published 2001

ISBN 0 7176 2016 6

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the copyright owner.

SUMMARY

General practitioner morbidity recording schemes may offer an opportunity to describe the frequencies of a wide range of conditions according to occupation.

The aim of the study was to examine the feasibility of assembling information nationally, on the frequencies and distribution of ill-health presenting to general practitioners, in relation to occupation.

Schemes were identified by literature search and discussions with expert personnel, including the HSE. Information was requested, by structured interview, from the scheme owners and operators, by meetings, telephone and post. Relevant peer review papers were also studied for details of methods.

In none of the Schemes is occupation routinely recorded, and such information would have to be collected by an additional procedure.

Together the four main ongoing Schemes can provide a population of sufficient size to detect differences in frequency of ill-health in most occupations. Statistical power to detect rare effects would be limited in the smallest occupational groups.

The criteria for data entry are broadly similar in the Schemes, and differences in completeness of entry at the individual doctor or practice level are likely to be more important than differences in criteria.

All Schemes monitor completeness of recording, in various ways, and feed the results back to the practices, so there is active encouragement of the quality of recording.

All Schemes use Read codes for diagnostic classification.

The Schemes differ in their use of independent ethics advisory committees. The main ethical concern is that there should be no possibility of identification of individual subjects. In all schemes, identities are coded in the practices before transmission of the clinical record. One concern is that adding occupation to the clinical record might enable identification of individuals.

All the Schemes are willing to discuss the possibilities for a combined study, subject to practical and ethical considerations.

A possible design for a pilot study is suggested.

CONTENTS

1. INTRODUCTION

1.1 NEED

The Health and Safety Executive seeks information on the frequency of occupational diseases, and their trends, and the effect of occupation on health. Specifically but not exclusively, the Health and Safety Commission Document 'Revitalising Health and Safety' {1} states a target to reduce the incidence rate of cases of work-related ill-health by 20% by 2010. This improvement will need to be measured.

Systematic information from employers is not readily available, and compensation scheme figures are open to selectivity, and influenced by compensation rules. NHS hospital records include inadequate occupational information. Several reporting schemes for specific occupational diseases are in operation and providing useful information, but are acknowledged to be subject to incomplete reporting, and diagnostic inconsistency.

An important concern also now is the effect of occupation on health in general, short of overt diagnosable occupational diseases, and schemes relying on recognition of the latter will miss less overt occupational health effects. HSE studies of self-reported occupational ill-health have suggested high frequencies of some disorders{2}. Another source of information on occupational health would usefully complement the existing schemes.

1.2 GENERAL PRACTITIONER MORBIDITY RECORDING SCHEMES

General Practitioner morbidity recording schemes offer an opportunity to describe the frequencies of a wide range of conditions according to occupation. Several schemes are known to be active, but are operated independently. The compatibility of information from them is not reported, nor the feasibility of obtaining information on the frequencies of diseases according to occupation. It is recognised that none of the schemes at present routinely obtain occupational histories systematically, and that this would be needed.

1.3 AIM AND OBJECTIVES OF THE STUDY

The aim of the proposed study is to examine the feasibility of assembling information nationally, on the frequencies and distribution of ill-health presenting to general practitioners, in relation to occupation. Specifically, the objectives are;

1. Identification of the major schemes in operation in the UK

2. Description of their modes of operation and scope, and, if possible, type of GP practice involved.

3. Description of patient inclusion criteria, methods of diagnostic classification, statistical summaries, ethical principles and procedures, restrictions and approving bodies

4. Discussion of the feasibility, benefits, limitations and costs of bringing together in some way the information from the schemes

5. If indicated, summarise a specification for work towards developing one or more of the schemes for the intended purpose

2. METHODS

Schemes were identified by literature search and discussions with expert personnel, including the HSE. Further investigations were confined to large national schemes with general rather than narrow disease interests. Information was requested, by structured interview, from the scheme owners and operators, by meetings, telephone and post, on current operations and relevant possibilities. Relevant peer review papers were also studied for details of methods. Methods for obtaining the occupational information were discussed, and conversations held with personnel familiar with the relevant research ethics committees on the acceptability of various methods for ensuring confidentiality.

The questions asked included; how might the data from the various sources be used to give an optimal summary of the frequencies of ill health according to occupation; what size of study population would be desirable to provide a sensitive indicator of occupational differences in ill health; what procedures would be feasible for matching occupational and ill health data; what procedures would be appropriate for ensuring confidentiality of these new data; what methods of obtaining occupational information would be feasible; what are the current procedures for preserving confidentiality; how are the data analysed and reported; to what extent are the results from the different sources comparable; the structure of systems for data gathering and transfer; what validation checks for diagnosis are conducted; how is ill health recorded, and what classification systems are used.

3. RESULTS

3.1 THE SCHEMES

Each scheme is based on one or more types of commercially available computer software for general practice medical records. The choice of software lies with the GPs. Details of the schemes are to be found in the appendix. The main schemes are as follows.

The General Practice Research Database (GPRD) managed by the Medicines Control Agency of the Department of Health{3}. Originally the research part of the Value Added Medical Database (VAMP) {4}, it was given to the government when Reuters purchased VAMP. It uses In-Practice Systems VM6 and Vision medical records computing software.

Mediplus UK Primary care Database (UKPCD), run by Intercontinental Medical Services (IMS). It is based on theTorex Meditel computer system.

The Doctors' Independent Network (DIN), a learned society and registered charity, based on Meditel and VampVision systems.

The GP Weekly Returns Service (GPWRS) of the Royal College of General Practitioners Research Unit {5}, funded by the department of Health. It is applied to Torex Meditel, and EMIS software.

A series of National Morbidity Surveys, based on mostly the same practices as the GPWRS, conducted jointly by the then OPCS and the RCGP Research Unit, from 1971 on. They were funded by the Department of Health. e.g. {6}, {7}. The last survey was in 1990-1 {8}. Occupational and other socio-economic information was obtained by special survey, and the methodology for this study may usefully inform the design of future studies {9}.

The Continuous Morbidity Recording Scheme (CMRS) operated by the NHS Common Services Agency in Scotland. The scheme is based on practices using G-Pass computer software, the use of which is encouraged by the NHS in Scotland.

Primary Care Information Services (PRIMIS) is operated by the Division of Primary Care, University of Nottingham, based usually on Torex Meditel or EMIS Systems. Its main interest is in quality of primary care. Data extraction and analysis is topic-based, arising from local initiatives (e.g. from Primary Care Groups).

3.2 POPULATION SIZES

Table 1 shows the numbers of practices (operating to acceptable quality standards) included in each scheme, and approximately estimated numbers of patients.

Table 1
Size of schemes and location of practices

Scheme	Practices	Current Patients (millions)	Area
GPRD	370	2.7	All UK
UK PCD (Mediplus)	125	1.0	All UK
DIN	100 selected for regular study. 300 in total.	1.1 (3.3 in total)	All mainland UK
CMRS	74	0.5	Mostly Scotland
GP WRS	72	0.6	England and Wales
National Morbidity Survey (4th)	60	0.4	England and Wales
PRIMIS	Varies by study		England (central)

3.3 STATISTICAL POWER

HSE provided information from the Labour Force Survey on the distribution of occupations of employees by 3-digit (minor group) codes in a sample of households for recent years, up to 1997-8. The Office of National Statistics, who conducted the studies, provided information, from the Spring 2000 survey, on the numbers surveyed (59402 households, results for 144,000 individuals over 16, 55,000 employees (38%), 8,000 self-employed (6%)). Assuming relative constancy, these figures were used to make approximate predictions of the numbers of general practice patients on average likely to be found in each occupation.

Examination of the resulting predicted distribution of occupational group sizes for employees (the self-employed are not included), indicates that of 371 occupations, 62 (17%) include less than 100 individuals per million adult population, 26 (7%) less than 50. The latter occupations are listed in table 2. Some of them could well be of interest from the point of view of their occupational health. Since there will be limited statistical power to detect differences between small groups, we suggest that these data provide an argument for maximising the size of the study population within practical constraints.

The other 83% of the occupations were adequately, or in many cases generously, represented.

Thus it appears that a total adult population of, say, three million adult patients would provide 371 occupational groups (not counting the self-employed), 93% of which include more than 150 individuals.

Assuming that three quarters of general practice patients are aged 16 or over, then a practice population of 4 million would provide an adult population of three million, and an employed population of about 1.1 million.

Table 2
Approximate predictions of frequencies of occupations: those including fewer than 50 people per million employed

Occupation	Number of individuals per million
Officers in non-UK armed forces	36
Prison principal officers and above	31
Senior customs and excise etc. officers	12
Barristers and advocates	32
Marine, insurance and other surveyors	44
Driving instructors (excluding HGV)	21
Tracers, drawing office assistants	25
Electrical production fitters	27
Barbenders, steel fixers	35
Music instrument makers, piano tuners	11
Scrap dealers, etc.	23
Tobacco process operatives	39
Tannery production operatives	39
Preparatory fibre processors	40
Winders, reelers	41
Metal drawers	23
Annealers, hardeners, etc. (metal)	46
Shot blasters	47
Bus inspectors	26
Bus conductors	18
Shunters and point operatives	20
Washers etc. in mines and quarries	33
Mine quarry workers (non-coal)	18
Fishing and related workers	21
Paviours and kerb layers	28
Slingers	26
Window cleaners	34

3.4 MODES OF OPERATION OF THE SCHEMES

Operating features common to all schemes include;

- The scheme operator provides information, on GP consultations, to GPs and to clients (government departments, pharmaceutical companies, researchers, community health bodies).

- In all schemes, to provide the data, the operator recruits and retains in the scheme general practices using certain clinical records software. Care is taken by the larger schemes to seek a representative sample of practices in relation to social levels and geographical location, though some of the smaller schemes are in the process of expanding from an initially limited area.

- The GP agrees to record clinical information according to defined quality standards. All the Schemes give advice, training, feedback on quality, and data summaries.

- All schemes use the Read Codes classification system for clinical information {10} These can be cross referenced to ICD codes {11}.

- Practice staff classify and anonymise the clinical records, and send selected fields to the scheme operator (automated in some Schemes).

- The scheme operator examines the quality of the data submitted, and provides feedback and training to practice personnel.

- The scheme operator analyses the data according to client requirements, or provides anonymised datasets to the client. GPRD licences approved researchers to examine the database online (DIN is planning to do this).

- Though frequency of data transmission varies, with one exception all systems permit the analysis of data by the summary time periods of interest for comparisons with occupation. The exception is the GPWRS, in which the weekly data is summarised automatically within the practices, and the ability to link to individual clinical records is not retained.

- In none of the schemes is the GP required to enter occupational information systematically, nor does the form provide a field for this, though the Read Codes classification provides for occupation. Examination of the summary data from one of the schemes indicated that occupation is very rarely reported in the narrative record.

Features which differ (significantly for present purposes) include;

- The criteria for making a clinical entry (though in practice this may not make much difference).

- Some scheme operators pay the GPs a fee per patient or record; others provide support in other ways, conditional on meeting quality standards.

- Geographical location of practices; for example English or Scottish predominance.

- Changes to systems; GPRD has recently redesigned its database, and the new system has not been in operation for many months; The G-Pass software on which CMRS is based

has also recently been redesigned, and CMRS reports are now being produced. VAMP6 and EMIS are continuous developments, while Vision is a new system.

- Most schemes use their own software to analyse the data. Some use versions of Miquest, the Department of Health system.

3.5 CRITERIA FOR MAKING A CLINICAL RECORD, AND QUALITY OF THE DATA

Doctors are exhorted in their training, and by the medical defence agencies to make adequate records of every consultation. The minimum criterion for inclusion of a record in any of these schemes is therefore that the doctor or practice staff think that a record was necessary. This applies, for example to the UKPCD scheme. At the other extreme, the GPRD scheme provides very detailed criteria for inclusion (see appendix). These appear to be a conscientious attempt to make explicit the criteria by which the doctor conventionally decides whether the consultation should be recorded. Other schemes require every face to face meeting to be recorded (CMRS and GPWRS for example). Details are given in the appendix.

It is likely that any biases induced by these differences will be smaller than the variation induced by errors in recording related to pressure of work, home consultations and other distractions.

Completeness of recording has been investigated by some schemes. For example review of hospital discharge summary information in the GPRD scheme indicated that the principal diagnoses were recorded correctly in 90% of computerised records. Other favourable results within the GPRD scheme {12-15} are summarised by Lawson et al {16}.

The Mediplus UKPCD scheme reports a 'very good' correlation between patient age in the database and national age across the range of patient ages, and in the split between males and females. The GP panel is broadly representative of the GP population of the UK, although there is some slight under-representation of smaller practices. In the Scheme, drug prescription frequencies have been shown to be comparable to national prescribing patterns, except for a 25% under-reporting of certain antibiotics commonly prescribed in home visits, presumably the result of the difficulties of making records after the event. Home visits conducted by a deputising service are unlikely to be recorded on the database. Main diagnosis or symptoms at other consultations appear to be reliably recorded (about 90%), but the results of hospital admissions and laboratory tests were less often entered (less than 50%) {17}.

In the 4th National Morbidity Study, a detailed analysis of quality indices was performed, showing low frequencies of wrong diagnosis recordings (only 4% wrong, 0.7% not reported) {8}.

In the CMRS in Scotland, a quality team visits the GP surgeries, and compares the computerised records with clinic and visit lists, identifies any omissions, examines the quality of recording, and advises accordingly.

In all the schemes, only data from practices reaching defined standards of recording are analysed. All schemes can provide information on the distribution of practices sizes, and comparisons with national norms.

3.6 ETHICAL AND CONFIDENTIALITY CONSTRAINTS

None of the schemes seek informed consent from the individuals for use of their data, though some practices make a general statement about participation in a scheme. The guiding ethical principle is that GPs and bona fide researchers may extract anonymous data from the clinical records for research purposes, providing there is no risk of linking clinical information to named individuals.

Guidance is provided by the Royal College of General Practitioners and Department of Health (not published), and most schemes have their own ethical advisory committee. The author has not been able to obtain documentation of these guidelines. In future either the Department of Health or the Multicentre Research Ethics Committee network may well be the appropriate approval bodies, but an approach has not yet been made by the author.

The ethical and confidentiality issues are;

- The identities of the participating practices are confidential to the related scheme operator, to avoid the linking of a rare diagnosis to a practice, and thus to an individual. There may also be an element of commercial confidentiality on this matter, the scheme operators protecting their sources.

- Before transmission of the clinical record, the identity of the patient must be represented only by a code for Practice and individual, the key being held only by the practice. Potentially identifying information in narrative parts of the record must be censored effectively.

- Schemes differ in whether they obtain the whole clinical record, including free text, or only approved fields; the fields common to all schemes include; Coded ID, year of birth (and month for young children), sex, date of consultation, diagnosis or symptoms, results of investigations or referrals, treatment. Postcode of residence is not allowed though two schemes allow a socio-economic class indicator.

- Opinions obtained informally from Scheme personnel, on inclusion of occupational information, varied; from serious concern that it would not be allowed because it may enable identification of individuals, to low concern because this is highly improbable, or because, in the case of DIN, the scheme is operated by GPs, who share anonymous information within a clinical confidentiality context. This is the most important ethical issue, and must be considered in any proposed study design.

- Any reports or publications must include only summary data, without identification of practice or individuals.

3.7 OCCUPATIONAL INFORMATION AND CONFIDENTIALITY; A SUGGESTION

Prompted by an approved arrangement used in France for conducting occupational mortality studies, we suggest that the anonymised occupational records be kept separate from the clinical records, and handled by a third party, who pools all the Practice occupational records, and censors occupations that are rare overall. The separate clinical records would then be obtained and matched only to the censored dataset, and after these have been linked, new IDs not showing Practice code would be substituted before statistical analysis.

3.8 PARTIES NECESSARY FOR A COLLABORATIVE STUDY

Collaboration between the various parties would be necessary for a study (unless a completely new self-contained study were to be established). They must include; one or more of the scheme operators (preferably several, for the sake of representativeness and statistical power), the GPs and their staff who participate in the schemes (known only to scheme personnel), the relevant ethical approval bodies, and a bona fide independent third party researcher competent in handling medical and occupational datasets. Coordination of the parties could be led by the HSE, say, or by a lead researcher.

Schemes with databases in which individual (though anonymous) clinical records can be distinguished include GPRD, DIN, UKPCD, CMRS and the National Morbidity Study. The latter, however is based on data up to 1984, and new data will not become available unless a new study is commissioned. The GPWRS, while well designed for its purpose, cannot identify individual records, and the PRIMIS scheme analyses data only project-by project.

3.9 OBTAINING OCCUPATIONAL INFORMATION

Since only the GPs know the identities of their patients, they or their staff (trained, or temporary assistants) will obtain occupational information. Since the information is not normally obtained for clinical purposes, and is intended for research, informed consent must be obtained. Staff must be trained in procedures, including coding, and information must be obtained even for subjects not consulting with their GP in the period. Occupational records would be allocated the same anonymous ID as for the clinical records Schemes, though the occupational record would not be attached to the clinical record, but sent directly to the independent third party researcher. The occupational and clinical records would not be linked until certain procedures to preserve anonymity had been performed.

3.10 TYPES OF POSSIBLE STUDY

A cross-sectional study of the relations between occupation and frequencies of diseases presenting to GPs would require occupational histories from several million patients to be obtained; a massive task. There would also be statistical difficulties in interpreting some of the unusual frequencies, because of chance variations and the large number of possible comparisons. This kind of approach would have the advantage of establishing a large database which could be used to study a wide range of issues, and could support prospective studies, but would be very expensive.

A case-control approach focussed on specific diagnoses would require fewer occupational histories, and would lessen the above statistical difficulty. In the course of time repeated case control studies would build up a database of randomly selected controls, with documentation of occupational histories, who might themselves provide a suitable and documented study population for examination of new questions without the need for further field studies.

3.11 DESIGN OF A STUDY

A possibility could be a collaborative study, a case/control approach, led by a steering group which would include members of GPRD, DIN, UKPCD and CMRS Schemes, with other experts, researchers and a lead organisation who would manage, write the proposal, reports, conduct work on behalf of the steering group and individual participants, etc.

A pilot study would be desirable, and might include these steps.

1. The lead organisation and steering group define the diagnostic categories of interest for the selection of cases and controls. It would be helpful for the lead organisation to analyse at least one of the databases to refine the definitions. For example request GPRD for a licence to analyse part of the database. The same could be done with the DIN database when the system has been established.

2. The Schemes are asked to identify internally the IDs of cases and controls according to these criteria, but would not release them.

3. GPRD and the other Schemes would ask the GPs for further information about these subjects. An aim should be to avoid loading any one practice with more than a few subjects. The steering group would provide clerical or nurse support for any work involved at the practices, but they would be acting on the GPs behalf. The GPs would be responsible (sign letters etc), and would give the work their approval.

4. The trained clerical person or nurse would go round the country liaising with Practice staff and doing the work of requesting the occupational information from the cases and controls. Informed consent would be requested, for use of the data, with assurances on confidentiality. The person would be acting in this context on behalf of the GP and practice staff. Alternatively a GP might prefer his own staff to conduct the work, possibly subject to funding assistance.

5. The clerical person or nurse, on behalf of the GP, would anonymise the occupational histories, using the ID code the GP uses for the Scheme, and would send to an independent research organisation (part of the team, could be the lead organisation), which would censor any rare occupations (less than, say, 50 individuals), to avoid any possibility of identifying an individual from an occupation. Pooling would be done nationally rather than by practice, otherwise too many occupations would be censored.

6. The independent researchers would request from the Schemes the case/control status of the individuals in the censored dataset, together with some non-clinical personal details (e.g. y.o.b., sex, maybe smoking), combine with the occupational data, assign new IDs to obscure the Practice ID, and analyse the results. The Schemes would also be asked to provide non-sensitive summary information on the non-responders (or the anonymised data). Members of the steering group would comment on the draft report, and co-author any reports. The lead researchers would make sure that commercially sensitive information was treated in confidence, between the collaborating parties, as well as externally.

7. Factors influencing response would be noted, together with the plausibility of the associations of disease with occupation. Progression to a full study would be determined by these results.

3.12 CONCLUSIONS

Together the four main ongoing schemes can provide a population of sufficient size to detect differences in frequency of ill-health in most occupations. Statistical power to detect rare effects would be limited in the smallest occupational groups.

In general the history of the Schemes indicates a tendency to change operation or ownership from time to time, and this is an additional reason to include several schemes in any plans to conduct possibly repeated studies.

The criteria for data entry are broadly similar in the Schemes, and differences in completeness of entry at the individual doctor or practice level are likely to be more important than differences in criteria.

All Schemes monitor completeness of recording, in various ways, and feed the results back to the practices, so there is active encouragement of the quality of recording.

All schemes use Read codes for diagnostic classification.

The schemes differ in their use of independent ethics advisory committees. The main ethical concern is that there should be no possibility of identification of individual subjects. In all schemes, identities are coded in the practices before transmission of the clinical record. One concern is that adding occupation to the clinical record might enable identification of individuals.

Personnel in all the schemes are willing to discuss the possibilities for a combined study, subject to practical and ethical considerations.

A possible design of a pilot study is suggested.

3.13 ACKNOWLEDGEMENTS

I thank the Scheme personnel for their willing cooperation, and the Health and Safety Executive for support for this study.

4. REFERENCES

1. Health and Safety Commission. Revitalising health and safety. London. Department of the Environment, Transport and Regions. 2000; OSCSG0390.

2. Jones, J.R., Hodgson,J.T., Clegg,T.A, Elliot,R.C. Self-reported work-related illness in 1995. Results from a household survey. London. HSE Books. 1998; ISBN 0 7176 1509 X.

3. Yamey G. Medicines Control Agency takes over GP research database. British Medical Journal. 1999; 319: 1153.

4. Mann RD. Electronic multipurpose database. Pharmaceutical Times. 1993; (June).

5. Fleming, D.M., Ross, A.M., Cross, K.W. et al. The weekly returns service. Annual report 1999. Birmingham. Birmingham Research Unit of the Royal College of General Practitioners. 2000.

6. Fleming DM. The measurement of morbidity in general practice. Journal of Epidemiology and Community Health. 1991; 45: 180-3.

7. Fleming, D.M., Crombie, D.L., and Ross, A.M. Weekly returns service. Report for 1996. 54, Lordswood Road, Harborne, Birmingham B17 9DB. Birmingham Research Unit of the Royal College of General Practitioners. 1997.

8. Office for Public Censuses and Surveys. Morbidity statistics from general practice. HMSO. 1995; ISBN 0 11 6916109.

9. Fleming DM, McCormick A, and Charlton J. The capture of socio-economic data in general practice. British Journal of General Practice. 1996; 46: 217-20.

10. NHS Information Authority. The clinical term version 3 (the Read codes). Loughborough. NHS Information Authority. 2000; 1999-IA-166 v1.0.

11. Saint Yves IF. The Read clinical classification. Health Bulletin Edinburgh. 1992; 50: 422-7.

12. Nazareth I, King M, Haines A, Rangel L, and Myers S. Accuracy of diagnosis of psychosis on general practice computer system. British Medical Journal. 1993; 307: 32-4.

13. Jick H, Jick SS, and Derby LE. Validation of information recorded on general practice based computerised data resource in the United Kingdom. British Medical Journal. 1991; 302: 766-8.

14. Jick H, Terris BZ, Derby LE, and Jick SS. Further validation of information recorded on a general practitioner based computerised data resource in the United Kingdom. Pharmacoepidemiology and Drug Safety. 1992; 13: 47-9.

15. van Staa T-P and Abenhaim L. The quality of information recorded on a UK database of primary care records: a study of hospitalisations due to hypoglycaemia and other conditions. Pharmacoepidemiology and Drug Safety. 1994; 31: 5-21.

16. Lawson DH, Sherman V, and Hollowell J. The general practice research database. QJM: Monthly Journal of the Association of Physicians. 1998; 91: 445-52.

17. Lawrenson RA, Coles G, Walton K, and Farmer RDT. Characteristics of practices contributing to the Mediplus database and the implications for its use in epidemiological research. Journal of informatics in primary care. 1998; 14-8.

5. APPENDIX

GP record data extraction schemes

Number
1

Scheme name
General Practitioner Research Database (GPRD)

Scheme type
Data extraction

Supplier
Medicines Control Agency (MCA)

Funding agencies or clients
Self-funding, clients; DoH, researchers, drug companies

Contact Name
Liz Morrison

Address
Medicines Control Agency, 15th Floor, Market Towers, 1 Nine Ems Lane, London SW8 5NQ

Phone Number	Fax Number	Email Address	Website
(0207) 2730206	(0207) 2730041	liz.morrison@gprd.com	www.gprd.com

Number of practices
370

Location of practices
UK

21 March 2001

Number of patients (millions)
2.7

Data transfer
Software in GPs surgery puts patient records, with identity coded, on to floppy disc, which is posted to GPRD. At start, whole records are transferred. Subsequently, six-weekly, only changes in record since last download.

Data interrogation

Data entry criteria
All significant morbidity events; 1) All events resulting in a hospitalisation or referral to any specialist, and the outcome of the referral (diagnosis, procedure, NAD etc). Any significant test results must be recorded. 2) All events resulting in a prescription or withdrawal of a drug or other treatment. This includes the indication for acute treatments, the original indication of every repeated treatment, and indication for any change in, or addition to, medication. All adverse reactions to drugs must be recorded. 3) Other events which the patient will consult with, on more than one occasion, or which the GP will require to be reminded of at a later date, including childhood diseases, pregnancy, when a patient has been asked to consult again after a period of time, etc. Note: a record must be made of those significant events identified by any member of the health team, whether the consultation occurred in the surgery, at a visit, or over the telephone. All diagnoses and procedures communicated to the GP as result of a hospital or other specialist visit (inpatient, outpatient or in an Accident and Emergency Unit) must be recorded when the GP is informed.

Classification of consultations
Not, but could do so after a fashion from complete clinical record (not equivalent to National Morbidity Surveys or CMRS

Classification of diagnoses
Read Codes. Previously OXMIS, from 1987. GPRD can now translate to Medra, MCAs own dictionary, but GPs still record using Read codes.

Data obtained
Coded ID, coded practice ID, YoB, sex, marital status,family number, date of event, diagnosis or symptoms, treatment, referrals, test results, deaths (sometimes cause)

Database details
Was flat files, just translated into database (Redbrick, Informix). Can be interrogated by Business Objects Software, then statistical packages used if needed.

Ethical procedures
Procedures and protocols reviewed by Ethics Advisory Board. IDs coded in GP surgery. Practice staff put double slash before all names and other identifiers in narrative record, and software censors record transmission record. GPRD further encrypts record to ensure absence of identifiers, and encodes practice number.

Associated data extraction/research schemes
NA

Associated GP records software
Based only on In-Practice Systems GP software, comprising the old VAMP system (VM6), and new, windows based 'Vision' software. Hoping to bring EMIS systems in.

Payments to GPs
25p per patient per year, plus training, guidance, quality checks and feedback.

Quality procedures
Checks on data received. Completeness, consistency, comparison with norms, reports to GPs, guidance. Only data from practices achieving defined quality standards are admitted to the database.

Services to researchers
On-line access to the database can be made available to trained or accredited research organisations, for specified regular access or one-off studies.

Remarks
Issue of ethical constraints on obtaining occupational information needs to be explored.

21 March 2001

Number
2

Scheme name
Mediplus

Scheme type
Data extraction

Supplier
Intercontinental Medical Statistics; IMS-Health

Funding agencies or clients
Drug companies, gov. depts.,MCA, researchers

Contact Name
Mr Peter Stephens, Ms Kruti Popat,

Address
107 Marsh Road, Pinner, Middlesex, HA5 5HQ

Phone Number	Fax Number	Email Address	Website
(0208) 7233400	(0208) 8688722	pstephens@uk.imshealth.c	

Number of practices
125 practices, 561 GPs (actually 800, but not all up to quality standards yet. Have records from 1980, 81.

Location of practices
UK. Selected to represent locality. Age sex profiles match gen. pop. Not allowed post code.

21 March 2001

Number of patients (millions)
1.0

Data transfer
Software in GPs system dials up Mediplus by dedicated ISDN line, and transmits anonymised, encypted. individual patient records. Is encrypted again in Meditel system, and individual records are updated as further events are recorded. Own software extracts data. Analyse data routinely using own software, plus SAS etc for special needs.

Data interrogation

Data entry criteria
At GPs initiative. Other practice personnel may enter data under GPs name. In practice most consultations leading to a prescription are entered, (the prescriptions are made using the system) except for home visits and cooperative clinics. 40% under reporting for Ampicillin analogues (commonest treatment prescribed at home visits). Hospital referrals recorded rather variably, also deaths.

Classification of consultations
No

Classification of diagnoses
Read Codes. Can translate into ICD 10. Codes

Data obtained
Age, sex, date of event, diagnosis, results of clinical tests, prescriptions, ID no.(anon), Practice ID, GP ID, dispensing practice or not.

Database details

Ethical procedures
Own Scientific and Ethics Committee. Follows BMA/RCGP guidelines. Get selected data from anonymised individual clinical records

Associated data extraction/research schemes

Associated GP records software
Based on selected Meditel user practices only. Contract is with GPs.

Payments to GPs
No. GPs get help, advice with reports, on coding, testing of templates, and support, and if achieving quality standards, points towards purchase of practice equipment

Quality procedures
Checks of data received, comparison of frequencies with national patterns of demographics and prescribing patterns. Monitor 11 quality parameters, on data completeness and patterns, provides feedback to GPs on quality. No filtering of admissable codes or ranges, known oddities in data, not excluded, but become apparent as unlikely values during analyses. Then excluded from those analyses, if appropriate.

Services to researchers

Remarks
Issues; 1) would GPs record occupation, 2) adding occupation directly to the clinical record may breach confidentiality rules; 3) IMS would be reluctant to identify the practices, for the same reason. If occupation is transmitted, it would have to be with permission of subject. Even then it would have to be held by a separate organisation which does not sell on data. Main difference from GPRD scheme is that in Mediplus, GP links prescription with diagnosis, under problem heading. Mediplus collects data systematically on clinical tests.

Number
3

Scheme name
Continuous Morbidity Recording Scheme (CMRS)

Scheme type
Data extraction

Supplier
Information and Statistics Division (ISD), NHS Common Services Agency (CSD), Scottish Executive.

Funding agencies or clients
Scottish Executive

Contact Name
Dr Jim Chalmers

Address
c/o C118, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ

Phone Number	Fax Number	Email Address	Website
(0131) 5518662		jim.chalmers@isd.csa.scot	Scottishhealth/ISD/Search CMR

Number of practices
75

Location of practices
Scotland (mostly). Fife Health Board adds some practices.

21 March 2001

Number of patients (millions)
0.5

Data transfer
In the clinic, GP has paper-based list of patients, made up by receptionist, with name, DoB, sex, practice code, date of consultation, space for morbidity, "new", "recurrent", "chronic" problem. Subsequently receptionist codes and punches the information into the practice computer on the G-Pass system, whicg also contains a unique identifying number. Once a month the receptionist dowmloads the information on to a floppy disc, and posts it to ISD. The file is a simple flat file. The only identification is the unique identifying number, the key being held only in the GPs clinical records.

Data interrogation

Data entry criteria
Every face to face doctor-patient consultation

Classification of consultations
First time problem has presented; recurrence of previously resolved problem; or persistent problem.

Classification of diagnoses
Read Codes

Data obtained
Patient ID, age, sex, postcode of residence, diagnosis, date of consultation, first time/recurrence/persistent problem code.

Database details

21 March 2001

Ethical procedures

Associated data extraction/research schemes

Associated GP records software
G-Pass

Payments to GPs
£4-5000 per year in total

Quality procedures
Checks at practices (40% so far); trained reception staff record all possible contacts, and these are compared with the number which have been recorded in the CMR way on the practice computer. In addition, for each practice, 80 patients' case notes are selected at random, scrutinised by the visiting team and ascribed Read codes by the visiting team. These codes are then compared to the codes which the practice has recorded. If the codes are wrong, then the practice is informed. If there is a serious coding problem, retraining of staff is arranged.

Services to researchers

Remarks
No particular ethical difficulties in asking for occupation, considering that identification of individuals from occupation is unlikely. GPs themselves would probably not be able to make the time to record occupation, but the practice nurse could. Or, the patient could be asked to complete a form themselves, either when attending or by post. We noted that informed consent could be obtained through this route. Probably extra funding would be required for this.

21 March 2001

Number
4

Scheme name
GP weekly returns scheme (GPWRS)

Scheme type
Data extraction

Supplier
Royal College of General Practitioners Birmingham Research Unit

Funding agencies or clients
DoH

Contact Name
Dr Douglas Fleming

Address
Lordswood House, 54, Lordswood Road, Harborne, Birmingham B17 9DB

Phone Number	Fax Number	Email Address	Website
(0121) 4261125	(0121) 4282084		

Number of practices
72

Location of practices
England and Wales

21 March 2001

Number of patients (millions)
0.6

Data transfer
Software in GP's system sends summary information automatically, twice a week, to GPWRS HQ

Data interrogation

Data entry criteria
Every face-to face doctor consultation

Classification of consultations
New problem, new spell, and recurrent or chronic conditions

Classification of diagnoses
Read Codes

Data obtained
Patient ID, year of birth, sex, dates,diagnosis, treatment, investigation, referrals, deaths.

Database details

Ethical procedures
Integral DoH procedures

Associated data extraction/research schemes

Associated GP records software
Meditel, EMIS

Payments to GPs
40p per registered patient per year

Quality procedures
Detailed checks on data received. See also National Morbidity Study

Services to researchers
Regular reports to DH

Remarks
Individual records not obtained

Number
5

Scheme name
National Morbidity Study

Scheme type
Series of one-off special studies. Data collection, extraction and analysis

Supplier
ONS, RCGP Research Unit, DoH

Funding agencies or clients
DoH

Contact Name
Dr Douglas Fleming

Address
Lordswood House, 54, Lordswood Road, Harborne, Birmingham B17 9DB

Phone Number	Fax Number	Email Address	Website
(0121) 4261125	(0121) 4282084		

Number of practices
60 last survey 1991/2

Location of practices
England and Wales

21 March 2001

Number of patients (millions)
0.5

Data transfer
Whole anonymised coded clinical record obtained in special cross-sectional exercise

Data interrogation

Data entry criteria
Every face-to-face consultation with doctor.

Classification of consultations
First ever, New or Recurrent, Ongoing Consultation

Classification of diagnoses
Read Codes

Data obtained
Full coded clinical record. Socio-economic data including occupation obtained by special surveys (last in 1991/2) (practice staff).

Database details

Ethical procedures
DoH procedures

Associated data extraction/research schemes

Associated GP records software
Meditel, EMIS

Payments to GPs
£2.19 per patient per year

Quality procedures
ONS team visited practices, inspected 100 records with the GP. Reported in detail in OPCS report (McCormick et al 1995)

Services to researchers
ONS makes information available to researchers

Remarks
1991/2 data base will be updated if a new cross-sectional study is commissioned

Number
6

Scheme name
PRIMIS (Primary care Information Services), previously Collection of Health Data from General Practice (CHDGP)

Scheme type
Data extraction

Supplier
Division of General Practice, University of Nottingham

Funding agencies or clients
NHS Information Authority, for 3 years.

Contact Name
Ms Sheila Teasdale

Address
Division of General Practice, University of Nottingham 14th Floor, Tower Building, University Park, Nottingham, NG7 2RD

Phone Number	Fax Number	Email Address	Website
(0115) 8466420	(0115) 8466432	sheila.teasdale@nottingha	

Number of practices
Varies according to project

Location of practices
England

21 March 2001

Number of patients (millions)

Data transfer
Use Miquest to extract information from GP records, mostly Meditel also Vision and Micromedic. Most other GP system suppliers are preparing MIQUEST interpreters or EMIS. Work is topic based, arising from local initiatives (eg from Primary Care Groups).

Data interrogation

Data entry criteria
Either; all morbidities at all patient consultations, or; recording data on a core set of morbidities

Classification of consultations

Classification of diagnoses
Read codes

Data obtained
Sex, year of birth, encrypted patient number)DIN doctors hold the encryption key, practice identity number

Database details

Ethical procedures
Follows standards laid down by the Committee on Standards of Data Extraction of the GPC.

Associated data extraction/research schemes

Associated GP records software
None, but often Meditel or Emis systems.

Payments to GPs
No

Quality procedures
Consistency checks on data extracted. List size vs. volume of notes recorded; list size vs. volume of drugs issues; percentage of drugs linked to coded reason for prescribing; proportion of acute to repeat prescriptions; visits and drugs prescribed entered on computer; proportion of 'firm' diagnoses vs. symptoms or signs; incidence of 'marker' Read codes. Software also provided for GPs to audit their own performance.

Services to researchers

Remarks
www.primis.nottingham.ac.uk Oriented towards training and support to local facilitators, quality of practice data management and computing. Projects address specific questions. Data not sold on.

21 March 2001

Number
7

Scheme name
Doctor Independent Network (DIN)

Scheme type
Data extraction

Supplier
DIN is a doctor's learned society, a registered charity.

Funding agencies or clients
NHS Executive, GP subscribers (members), drug companies, researchers.

Contact Name
Dr Paul Steventon

Address
Fitznells Manor Surgery, Chessington Road,Ewell, Surrey KT17 1TF

Phone Number	Fax Number	Email Address	Website
(0208) 3941481	(0208) 3939753	pauls@DINBOARD.demon.	

Number of practices
300 in total, 1400 GPs, a selected 100 practices supplying high quality data.

Location of practices
All UK mainland, Scotland slightly under-represented, a few in Channel Islands and N. Ireland

Number of patients (millions)
Current number not available. Estimate based on 2400 patients per GP suggests 3.4 million overall, 1.1million for selected 100 practices.

Data transfer
Extraction software works on Torex Meditel System 5 and System 6000 (and developing for In-Practice Systems VAMP Vision) files in practices, extracts usually daily by modem, into DIN database. Uses Unix operating system. Transmits all changes to clinical records since last interrogation.

Data interrogation
Own design software.

Data entry criteria
All GP face-to face consultations, referrals, tests, hospital letters.

Classification of consultations
No

Classification of diagnoses
Read Codes

Data obtained
Sex, year of birth, CACI socio-economic code collection, encrypted patient number, doctor code, DIN doctors hold the encryption key, practice identity number (DIN-specific), diagnosis, symptoms, treatment, test results, hospital referrals.

Database details
Sculptor

Ethical procedures
Follows standards laid down by the Committee on Standards of Data Extraction of the GPC.

Associated data extraction/research schemes
CompuFile Ltd is an independent company which markets DIN data

Associated GP records software
Torex Meditel System 5 and 6000, developing to include VAMP Vision

Payments to GPs
Only in training, advice, software.

Quality procedures
Consistency checks on data extracted. List size vs. volume of notes recorded; list size vs. volume of drugs issued; ratios of acute to repeat prescriptions; percentage of drugs linked to coded reason for prescribing; proportion of acute to repeat prescriptions; visits and drugs prescribed entered on computer; proportion of 'firm' diagnoses vs. symptoms or signs; incidence of 'marker' Read codes. Software also provided for GPs to audit their own performance.

Services to researchers
Will provide anonymised database to bona fide researchers. DIN-Link provides summary data for clients. Developing website which will make database available directly to bona fide researchers.

Remarks
Oriented towards clinical audit and improvement, also provides data for drug companies and researchers. Dr Steventon enthusiastic about extending services and participating in possible future research.

Number
8

Scheme name
DIN-Link

Scheme type
Data extraction

Supplier
CompuFile Ltd

Funding agencies or clients
Pharmaceutical companies, Health economics companies, resaearch charities (eg British Thoracic Society)

Contact Name
Mrs Emma Watkin

Address
1 Tannery House, Tannery Lane, Send, Woking, Surrey, GU23 7EF

Phone Number	Fax Number	Email Address	Website
(0148) 3222848	(0148) 3211037	EmmaW@Compufile.co.uk	www.compufile.co.uk

Number of practices
A panel of 100 practices is selected from 180 available

Location of practices
All mainland UK, Scotland underrepresented

21 March 2001

Number of patients (millions)
Estimated at 2 million in the 180 practices which gives approximately 750,000 in the 100 practice panel

Data transfer
Based on Torex Meditel Software in practices, provided by DIN to GPs, transmits anonymised patient records nightly by modem; activated by DIN.

Data interrogation
Database interrogated by 'Overview' software (own design). Can export to Excel for additional analyses.

Data entry criteria
Letters from hospital may be under-represented as GPs do not always input details of letters received into patient records.

Classification of consultations
No

Classification of diagnoses
Read Codes

Data obtained
Individual clinical records. Practice ID, anonymised patient ID (allocated by DIN). Dates of registration, deregistration, consultations, DoB, sex, Diagnosis, symptoms, treatment, tests, referrals.

Database details
Own design. Dates from 1988.

21 March 2001

Ethical procedures
Consultation with DIN. Anonymised patient records. Published data must be summarised across UK, without identifying practices.

Associated data extraction/research schemes
DIN

Associated GP records software
Torex Meditel

Payments to GPs
No. Practice software and audit data provided.

Quality procedures
Criteria for practice data entry to database; fundolding practices, range of sizes (single to 6 man pratices), representative geographica/population distribution, five years of consecutive data, without breaks in excess of five days. A minimum of 80% linkage is required, but this is currently approximately 90%.

Services to researchers
'Provision of "Overview" summaries, anonymised clinical records, licences to interrogate the database.

Remarks

21 March 2001

Printed and published by the Health and Safety Executive
C1 04/01

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from booksellers

HEALTH AND SAFETY INFORMATION
HSE InfoLine
Tel: 08701 545500
Fax: 02920 859260
e-mail: hseinformationservices@natbrit.com
or write to:
HSE Information Services
Caerphilly Business Park
Caerphilly CF83 3GG

HSE website: www.hse.gov.uk

CRR 340

£10.00

ISBN 0-7176-2016-6