

Research article

Open Access

Implications of the problem orientated medical record (POMR) for research using electronic GP databases: a comparison of the Doctors Independent Network Database (DIN) and the General Practice Research Database (GPRD)

Iain M Carey¹, Derek G Cook*¹, Stephen De Wilde¹, Stephen A Bremner¹, Nicky Richards², Steve Caine², David P Strachan¹ and Sean R Hilton¹

Address: ¹Department of Community Health Sciences, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE and ²CompuFile Ltd., Send, Surrey GU23 7EF

Email: Iain M Carey - sgjd450@sghms.ac.uk; Derek G Cook* - d.cook@sghms.ac.uk; Stephen De Wilde - sdewilde@sghms.ac.uk; Stephen A Bremner - sbremner@sghms.ac.uk; Nicky Richards - NickyR@Compufile.co.uk; Steve Caine - SteveC@Compufile.co.uk; David P Strachan - sgjd950@sghms.ac.uk; Sean R Hilton - shilton@sghms.ac.uk

* Corresponding author

Published: 30 September 2003

Received: 09 July 2003

BMC Family Practice 2003, **4**:14

Accepted: 30 September 2003

This article is available from: <http://www.biomedcentral.com/1471-2296/4/14>

© 2003 Carey et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The General Practice Research Database (GPRD) and Doctor's Independent Network Database (DIN), are large electronic primary care databases compiled in the UK during the 1990s. They provide a valuable resource for epidemiological and health services research. GPRD (based on VAMP) presents notes as a series of discrete episodes, whereas DIN is based on a system (MEDITEL) that used a Problem Orientated Medical Record (POMR) which links prescriptions to diagnostic problems. We have examined the implications for research of these different underlying philosophies.

Methods: Records of 40,183 children from 141 practices in DIN and 76,310 from 464 practices in GRPD who were followed to age 5 were used to compare the volume of recording of prescribing and diagnostic codes in the two databases. To assess the importance and additional value of the POMR within DIN, the appropriateness of diagnostic linking to skin emollient prescriptions was investigated.

Results: Variation between practices for both the number of days on which prescriptions were issued and diagnoses were recorded was marked in both databases. Mean number of "prescription days" during the first 5 years of life was similar in DIN (19.5) and in GPRD (19.8), but the average number of "diagnostic days" was lower in DIN (15.8) than in GPRD (22.9). Adjustment for linkage increased the average "diagnostic days" to 23.1 in DIN. 32.7% of emollient prescriptions in GPRD appeared with an eczema diagnosis on the same day compared to only 19.4% in DIN; however, 86.4% of prescriptions in DIN were linked to an earlier eczema diagnosis. More specifically 83% of emollient prescriptions appeared under a problem heading of eczema in the 121 practices that were using problem headings satisfactorily.

Conclusion: Prescribing records in DIN and GPRD are very similar, but the usage of diagnostic codes is more parsimonious in DIN because of its POMR structure. Period prevalence rates will be underestimated in DIN unless this structure is taken into account. The advantage of the POMR

is that in 121 of 141 practices using problem headings as intended, most prescriptions can be linked to a problem heading providing a specific reason for their issue.

Background

Primary care computing in the United Kingdom has a longer history than in any other country[1]. The design of the early general practice software systems (many of which continue in use in developed forms) grew out of the personal preferences for particular styles of medical record of the general practitioners who designed and used the systems. Of the two systems developed in the mid 1980s, Meditel System 5 software was heavily influenced by the concept of the Problem Orientated Medical Record (POMR)[2]. It was designed to present the medical record as a chain of intertwined but discrete problems, with prescriptions being linked to diagnoses under problem headings. In contrast VAMP Medical software presented the notes as a series of discrete episodes, essentially unconnected, an approach taken by a number of other software suppliers.

An early business model on the part of the two software suppliers (VAMP Health, now IPS, and Meditel, now Torex) envisaged the provision of free or subsidised computer hardware to general practitioners in exchange for their contribution of coded diagnostic and prescribing data to the supplier's database. Neither supplier now has any connection with the databases they created (the General Practice Research Database (GPRD) (VAMP) and the Doctors' Independent Network Database (DIN) (Meditel)), but both databases continue to collect data. These large scale databases built up during the 1990's provide a valuable resource for epidemiological and health services research. However, the implications for research of the different philosophies underlying the databases as well as their different system for coding diagnoses and prescribing (Oxford Medical Information System [OXMIS] and subsequently 5-Byte Read codes in GPRD, and 4-Byte Read codes in DIN) have been little studied. To do so is important for two reasons: first, to make best use of the large volumes of data collected during the 1990's; second, to inform the development of the electronic patient record, which thus far appears to have proceeded without formal comparisons of the different databases built up.

The problem orientation of DIN with linkage of prescriptions to problems enables, in principle, reasons for all prescriptions to be clearly identified. The absence of such linkage in GPRD means that contributors of data to the database are expected to code a specific reason for each prescription issued. While in theory this occurs for acute prescriptions, allowing linkage by date, only the original

indication for repeated treatment is required to be entered[3].

In this study we assess the importance of these differences by analysing birth cohorts in DIN and GPRD to: (i) compare the volume of recording of prescribing and diagnostic codes in DIN and GPRD (both overall and by examining variation between practices); (ii) examine the value of linkage for identifying explicitly the reason for prescribing. The latter is achieved by focussing on reasons for prescriptions of skin emollients.

Methods

Background to DIN and GPRD

We identified 142 UK general practices from the DIN database that were considered good quality data providers, based on a series of indicators which we have previously described[4]. Here we use data recorded between January 1989 and February 2002 – one practice is excluded from this paper because it did not contribute for a full 5-year period. The main data considered here consist of the amalgamation of two files – "Notes" and "Issues". The "Notes" file covers diagnoses, symptoms, administration, medical history, vaccinations and prescriptions other than repeats. The "Issues" file contains more detailed prescribing data on all prescriptions including repeats. DIN practices used Read Codes for recording both drugs and diagnoses.

An example of a record from DIN, ordered by event date, and including entries from both the "Notes" and "Issues" file for one child is given in Table 1. A diagnosis of eczema is made on 9/2/96 and a prescription for hydrocortisone cream issued. In this instance the diagnosis of eczema is the problem heading and is assigned Problem Number 5. On 2/4/96 and 24/6/96 emollient prescriptions are recorded under the same problem heading. Notice that more than one diagnosis can be placed under this heading if necessary. Thus if the child developed nappy rash later, the GP could choose to add this under the existing problem heading. The alternative would be to start a new heading.

The GPRD is well established in epidemiological research[5]. Here we consider data collected from 464 practices between 1989 and 1998 only. For much of this period, practices in GPRD used the OXMIS coding system, but since 1997 some practices have used Read Codes (Unified 5-Byte Version 2 Set) for diagnoses. PPA codes

Table 1: Example record from DIN illustrating linkage by problem number (Selected variables from the combined "Notes" and "Issues" files are ordered by event date. Type C indicates a chapter heading)

| Event | Seq | Probno | Type | Dosea | Dosei | Repeat | ReadRC | ReadRu |
|----------|-----|--------|------|-------|-------|--------|--------|----------------------------------|
| 25/12/95 | 8 | 4 | C | | | | 63.. | Birth details |
| 25/12/95 | 10 | 4 | N | | | | K4B1 | Normal delivery – occipitoant. |
| 25/12/95 | 12 | 4 | P | | | | 6363 | Baby BW = 10%-24% (2850–3149 g) |
| 25/12/95 | 14 | 4 | P | | | | 62PI | Breast fed |
| 08/01/96 | 2 | 1 | C | | | | 65.. | Infectious dis:prevent/control |
| 08/01/96 | 4 | 2 | C | | | | 9192 | Registered child surveillance |
| 08/01/96 | 6 | 3 | C | | | | 133D | Child of |
| 10/01/96 | 16 | 225 | A | | | | 9123 | Patient reg. form sent to FPC |
| 10/01/96 | 18 | 225 | A | | | | 919Z | Child surveillance admin.NOS |
| 09/02/96 | 20 | 5 | C | | | | L2.. | Dermatitis/eczemas |
| 09/02/96 | 21 | 5 | M | 15 | a2 | A | m411 | HYDROCORTISONE 1% cream 15 g |
| 09/02/96 | 22 | 2 | P | | | | 64D4 | Child 6 week exam. normal |
| 26/02/96 | 24 | 1 | P | | | | 6511 | First DTP (triple)+polio vacc. |
| 26/02/96 | 26 | 1 | P | | | | 657A | 1st haemophilus B vaccination |
| 25/03/96 | 28 | 1 | P | | | | 6512 | Second DTP (triple)+polio vacc |
| 25/03/96 | 30 | 1 | P | | | | 657B | 2nd haemophilus B vaccination |
| 02/04/96 | 37 | 5 | M | 100 | Up | A | m271 | OILATUM PLUS bath emollient |
| 22/04/96 | 38 | 1 | P | | | | 6513 | Third DTP (triple)+polio vacc. |
| 22/04/96 | 40 | 1 | P | | | | 657C | 3rd haemophilus B vaccination |
| 24/06/96 | 43 | 5 | M | 400 | — | A | m23k | SUDOCREM cream 400 g |
| 24/06/96 | 45 | 5 | M | 500 | Ap | A | m111 | AQUEOUS cream |
| 24/06/96 | 47 | 5 | M | 100 | Up | R | m271 | OILATUM PLUS bath emollient |
| 23/08/96 | 48 | 2 | P | | | | 64E4 | 8–9 month exam normal |
| 30/08/96 | 50 | 2 | P | | | | 64M5 | Child exam.: hearing |
| 24/02/97 | 52 | 1 | P | | | | 65M1 | Measles/mumps/rubella vaccn. |
| 14/07/97 | 54 | 2 | c | | | | 64F4 | 18 month exam. Normal |
| 30/03/99 | 56 | 2 | P | | | | 64H4 | 3.5 year exam. Normal |
| 30/03/99 | 58 | 2 | P | | | | 64N7 | Child exam.: testes |
| 01/04/99 | 60 | 6 | C | | | | 22G2 | O/E – parotid swelling |
| 05/05/99 | 62 | 7 | C | | | | H161 | Acute bronchitis |
| 05/05/99 | 63 | 7 | M | 100 | s3 | A | e3zk | AMOXYCILLIN 125 mg/5 mL s/f susp |
| 09/08/99 | 64 | 8 | C | | | | L15. | Impetigo |
| 09/08/99 | 65 | 8 | M | 100 | s4 | A | e22j | FLUCLOXACILLIN 125 mg/5 mL syrup |
| 15/10/99 | 66 | 9 | C | | | | F631 | Acute suppurative otitis media |
| 15/10/99 | 67 | 9 | M | 100 | s3 | A | e31k | AUGMENTIN 125/31 in 5 mL susp |
| 05/01/00 | 68 | 1 | P | | | | 65K4 | Booster DT(double)+polio vacc. |
| 05/01/00 | 70 | 1 | P | | | | 65MA | MMR booster |
| 02/08/00 | 72 | 1 | P | | | | 657E | First meningitis C vaccination |
| 02/08/00 | 74 | 1 | P | | | | 657I | Single meningitis C vaccinatn |

Key: Event Day of Event. **Seq** Sequence Number. This is a unique sequential identifier for each entry in the "Notes". It also provides the link between "Notes" and "Issues". **Probno** Problem Number. This identifies all entries linked together under the problem-orientated linkage. **Type** Note Type. Defines the entry (C = Current/Major Problem Heading, M = Medication, P = Prevention, A = Administration and so on) **Dosea** Dosage Amount. Refers to drugs entries in the "Issues" only. **Dosei** Dosage Instructions. Refers to drugs entries in the "Issues" only. **Repeat** Repeat / Acute Flag. Refers to drugs entries in the "Issues" only. **ReadRC** 4 Byte Read Code. **ReadRu** Rubric for Read Code.

were used for recording drugs, which were mapped to BNF chapters using a browser.

In preliminary work we compared all patients fully registered in DIN in 1998 (approximately 1 million subjects), to that of GPRD in the same year (based on Key Health Statistics[6]). We found the age-sex population structure of the two databases to be highly comparable to each other and National Statistics population estimates. However there were differences in geography, with DIN having 60.3% of its population in the South, compared to only 41.7% for GPRD, neither of which compare closely to the population estimate of 49.8%. It was important to adjust

for this difference when we compared the prevalence of ischaemic heart disease that is known to vary by region[4].

Defining birth cohorts in each database

We created a subset of each database that included only children who were "born into" the database, which we defined as those fully registered within 3 months of their date of birth[7]. In this report we further restrict to children who were continually registered up to age 5. Using birth cohorts in the analyses have a number of advantages: (i) all consultations should be recorded in 'real-time' and there would be no reliance on retrospective data; (ii) comparisons within and between databases

would be fair by having a fixed follow-up period; (iii) the data sets were of manageable size.

Comparisons of the volume of recording of diagnoses and prescribing in GPRD and DIN

To further investigate the extent to which DIN relies on linkage, the average number of prescriptions versus the average number of diagnoses (and symptoms) in the first 5 years of life was compared between practices within databases, and between databases. Read and OXMIS Codes for entries referring to administration, family history of disease, procedures and examinations were excluded. We made the pragmatic decision to refine the analysis by measuring diagnosis *days* and prescription *days* – that is number of days on which a diagnosis was recorded or a drug was prescribed. In DIN, a day on which a prescription was made was then defined as being *linked to a diagnosis* if the problem number it was associated with had appeared previously in the child's record with a diagnosis (or symptom).

Identifying DIN practices using linkage satisfactorily and assessing its appropriateness

In DIN, the percentage of prescriptions that link to a problem number is high (>80% at least for all practices, >90% for most) because it was one of the data quality indicators used to select practices, and data from within practices. However, this indicator of linkage does not guarantee a properly structured problem orientated medical record. For example, a practice could link all its prescriptions to a single problem number. Such a practice might be keeping excellent records, and the data would be useful for many purposes, but linkage of problem headings would be meaningless.

Three criteria based on diagnoses found in the problem headings were used to identify practices which were not using linkage in a way useful for research: (i) reliance on non-specific Read Codes for problem headings (11 practices had >20% 'level 1' Read codes); (ii) reliance on practice-specific Read Codes for problem headings for which we have no rubric (7 practices had >20% such codes) and; (iii) opening too few problem headings per person on average (2 further practices were clear outliers in this regard). This left 121 practices that we believed would be suitable for an analysis based on linkage to problem headings.

To assess the reliability and potential usefulness of linkage in DIN, the analysis focused on emollients (BNF chapter 13.2.1). Emollients were chosen as they are predominantly prescribed for a single condition (eczema), frequently diagnosed and prescribed for in general practice. Eczema was defined by the following Read Codes – L2.. (dermatitis/eczemas), L22. and all sub-codes (Sebor-

rhoeic dermatitis/eczema), L23. and all sub-codes (Atopic eczema/dermatitis), L24. and all sub-codes (Contact dermatitis/eczema), L25. and all subcodes (ingestion dermatitis), F5C4 (dermatitis of eyelid) and 2F13 (dry skin). We assessed emollient to eczema linkage in three ways in DIN – same day, linked by problem number anywhere in the record, and linked to a problem heading only.

Computing

Data handling at St George's was carried out using SAS version 8.1 (SAS Institute Inc., North Carolina, USA) running under UNIX on a Sun Microsystem dual processor with large scale rapid access storage using RAID technology.

Statistical methods

The paper is essentially descriptive, comparing the level of recording of various items within two large databases of tens of thousands of patients. Means, medians and inter-quartile ranges are used to characterise the distributions, many of which are skewed. Despite this skewness, means are well estimated based on large numbers of observations, and have the merit of relating directly to the total level of recording/prescribing. For this reason means were also used for studying variation in record lengths between practices. PROC MIXED in SAS was used to estimate the within practice and between practice components of variance within each database (with practice fitted as a random effect). The average percentage of variation between practices which was attributable to chance (sampling variation) was then calculated. Plots of the mean level of the number of days with diagnostic codes against the number of days on which prescriptions were issued were restricted to practices contributing more than 25 children to limit the role of sampling variation.

Results

111,621 children were deemed to be "born into" DIN, and 504,273 "born into" GPRD. We study here only those with 5 years continuous follow-up: 40,183 in DIN and 76,310 in GPRD.

Comparison of the birth cohorts

The main differences between the birth cohorts are highlighted in Table 2. Since the DIN database builds up from the early 90's, peaking in 1998 before practices started to switch to the newer System 6000 – the birth year with the most births who were followed for 5 years in DIN was 1994. (As we only consider DIN data recorded to February 2002 here, there are very few 1997 births). By contrast, GPRD data available to us only existed to 1998, so there are no births after 1993. The smaller range of birth years in GPRD also resulted in fewer children per practice (less than 100 in 69% of practices compared with in 37% of practices in DIN). There were no important differences in

Table 2: Comparison of the DIN and GPRD birth cohorts

| | | DIN (141 practices) | | GPRD (464 practices) | |
|---|-------|---------------------|-------------|----------------------|-------------|
| | | # children | % | # children | % |
| Total with 5 years follow up | | 40,183 | | 76,310 | |
| Sex | Boys | 20,575 | 51.2% | 38,904 | 51.0% |
| | Girls | 19,608 | 48.8% | 37,406 | 49.0% |
| Birth Year | 1989 | 511 | 1.3% | 13,772 | 18.0% |
| | 1990 | 3,003 | 7.5% | 18,946 | 24.8% |
| | 1991 | 4,421 | 11.0% | 17,532 | 23.0% |
| | 1992 | 5,412 | 13.5% | 14,646 | 19.2% |
| | 1993 | 6,802 | 16.9% | 11,414 | 15.0% |
| | 1994 | 7,410 | 18.4% | 0 | 0% |
| | 1995 | 6,730 | 16.8% | 0 | 0% |
| | 1996 | 5,373 | 13.4% | 0 | 0% |
| | 1997 | 521 | 1.3% | 0 | 0% |
| | | | # practices | % | # practices |
| Number of children contributed to the birth cohort by practice | | | | | |
| 0-100 | | 20 | 14.2% | 172 | 37.1% |
| 100-200 | | 32 | 22.7% | 146 | 31.5% |
| 200-300 | | 33 | 23.4% | 80 | 17.2% |
| 300-400 | | 22 | 15.6% | 38 | 8.2% |
| 400-500 | | 20 | 14.2% | 18 | 3.9% |
| >500 | | 14 | 9.9% | 10 | 2.1% |

the record length indicators between the DIN children born in 1989-93 (where we have GPRD data) and those in 1994-97 (Table 3). All subsequent analyses are thus based on the combined DIN dataset over all years.

The mean number of days per patient on which a drug was prescribed was similar between databases (19.5 in DIN vs. 19.8 in GPRD, Table 2), but mean days with a diagnostic code were lower in DIN (15.8 vs. 22.9 in GPRD). Including prescriptions that were linked to a diagnostic code raised the DIN mean to 23.1 days. The mean number of days per patient with either a drug prescribed or a diagnostic code recorded (or both) was slightly higher in GPRD than in DIN (27.4 v 26.0 days).

While differences in mean record lengths between databases were small after allowing for linkage, differences between practices were marked within both databases (Table 4). Thus the practice means for the average number of days in which a prescription was issued in the first 5 years of life had an inter-quartile range of 16.4 to 22.3 in DIN and 15.5 to 22.5 in GPRD (Table 4). Very little of the practice variation is due to chance variation - that is sampling variation arising from which individuals by chance were in a given practice. The percentage due to sampling variation is smaller in DIN due to the greater number of children per practice. The percentage for number of days

with diagnostic codes is particularly small in DIN because this figure is highly dependent on the way in which practices structure their records; the real differences between practices are thus correspondingly greater.

The inter-practice relationship between prescribing and recording of diagnostic codes is displayed in Figure 1, separately for each database. In order to limit the role of sampling error, only practices with at least 25 children are included in the plots. Generally high prescribing practices are also those which use the most diagnostic codes, however the correlation was much weaker in DIN ($r = 0.41$, Figure 1a) than in GPRD ($r = 0.81$ Figure 1c). Adjustment for linkage in DIN produced a similar correlation ($r = 0.85$, Figure 1b) to that seen in GPRD.

Linkage of emollients to eczema in DIN

More children had a prescription for an emollient in DIN (41.8%) than in GPRD (36.6%) during their first five years of life (Table 5). Although these children were more likely to have had an eczema diagnosis in DIN (91.7%) than in GPRD (80.3%), the diagnosis was less likely to appear on the same day as the emollient prescription in DIN than in GPRD (19.4% v 32.7%). However, in DIN 86.4% of all emollient prescriptions appeared under a problem heading which also included a diagnosis of

Table 3: Comparison of record length in the DIN and GPRD birth cohorts

| | DIN | | | | | GPRD | | |
|--|--------------------------------|--------------------------------|-------------------------------|---------------------|------------------|-----------------------------|---------------------|------------------|
| | 1989-93 births (n = 20,149) | 1994-97 births (n = 20,034) | 1989-1997 births (n = 40,183) | | | 1989-93 births (n = 76,310) | | |
| | Mean ^① | Mean ^① | Mean ^① | Median ^② | IQR ^③ | Mean ^① | Median ^② | IQR ^③ |
| No. of days with a prescription issued | 19.6 | 19.4 | 19.5 | 14 | 8-24 | 19.8 | 15 | 9-26 |
| No. of days with diagnostic codes | 15.8 | 15.7 | 15.8 | 13 | 8-21 | 22.9 | 20 | 13-30 |
| No. of days with diagnostic codes or drugs linked to existing diagnostic codes | 23.0 | 23.2 | 23.1 | 18 | 10-30 | - | - | - |
| No. of days with either diagnostic or drugs or both codes | 25.9 | 26.0 | 26.0 | 20 | 12-33 | 27.4 | 23 | 14-35 |

① – mean number of events in the first 5 years of life averaged across all individuals ② – median number of events in the first 5 years of life across all individuals ③ – lower and upper quartiles of number of events in the first 5 years of life across all individuals

Table 4: Practice variation in average record lengths in the DIN and GPRD birth cohorts

| | DIN (n = 141) | | | | GPRD (n = 464) | | | |
|--|---------------------|---------------------|---------------------|--|---------------------|---------------------|---------------------|--|
| | Lower quartile ① | Median ^① | Upper quartile ① | % variation between practices due to sampling variation ^② | Lower quartile ① | Median ^① | Upper quartile ① | % variation between practices due to sampling variation ^② |
| No. of days with a prescription issued | 16.4 | 18.5 | 22.3 | 5.2% | 15.5 | 18.4 | 22.5 | 8.3% |
| No. of days with diagnostic codes | 10.6 | 15.4 | 20.0 | 1.5% | 18.5 | 21.9 | 25.6 | 8.5% |
| No. of days with diagnostic codes or drugs linked to existing diagnostic codes | 18.8 | 22.9 | 27.3 | 4.8% | - | - | - | - |
| No. of days with either diagnostic or drugs or both codes | 21.1 | 25.6 | 30.8 | 4.8% | 22.5 | 25.7 | 30.7 | 9.9% |

① – the 25th, 50th and 75th percentiles of the practice mean number of events in the first 5 years of life ② – see statistical methods

eczema (whether as the problem heading or simply as one of several linked diagnoses) and 75.4% of all emollient prescriptions appeared under a problem heading of eczema.

However, there is marked variation between practices in the percentage of emollient prescriptions that link to a problem heading of eczema (Figure 2). It is apparent that those 20 practices identified a-priori as not using problem headings satisfactorily included all but two practices with poor emollient to eczema linkage (Figure 2). These two were readily explained by their reliance on Read codes ("2227" Rash Present and "L4ZZ" Skin disorders not otherwise specified) that did not fit our definition of eczema, but in all likelihood were used by the practices to represent it. After excluding these 20 practices, the percentage of all emollient prescriptions appearing under a problem heading of eczema rose to 83.5%.

An analysis of the problem headings under which emollient prescriptions were issued was carried out for the 121 practices using problem headings satisfactorily. A total of 14,938 problem headings were used under which a prescription for an emollient was issued. Of these problem headings 76.9% (11,480/14,938) were defined as eczema; 4.7% (709) were other skin conditions (e.g. "L15." Impetigo); 7.6% (1,135) were other skin symptoms (e.g. "2227" Rash Present); and 4.6% (689) were non-specific entries (which may well be eczema e.g. "L...;" Skin/subcutaneous tissue disease). Only 6.2% (925) of headings appeared to be unsuitable reasons for an emollient prescription.

Discussion

This report has illustrated some similarities and differences between two large-scale UK general practice computer databases. While both have similar levels of prescribing, problem orientated linkage in DIN results in

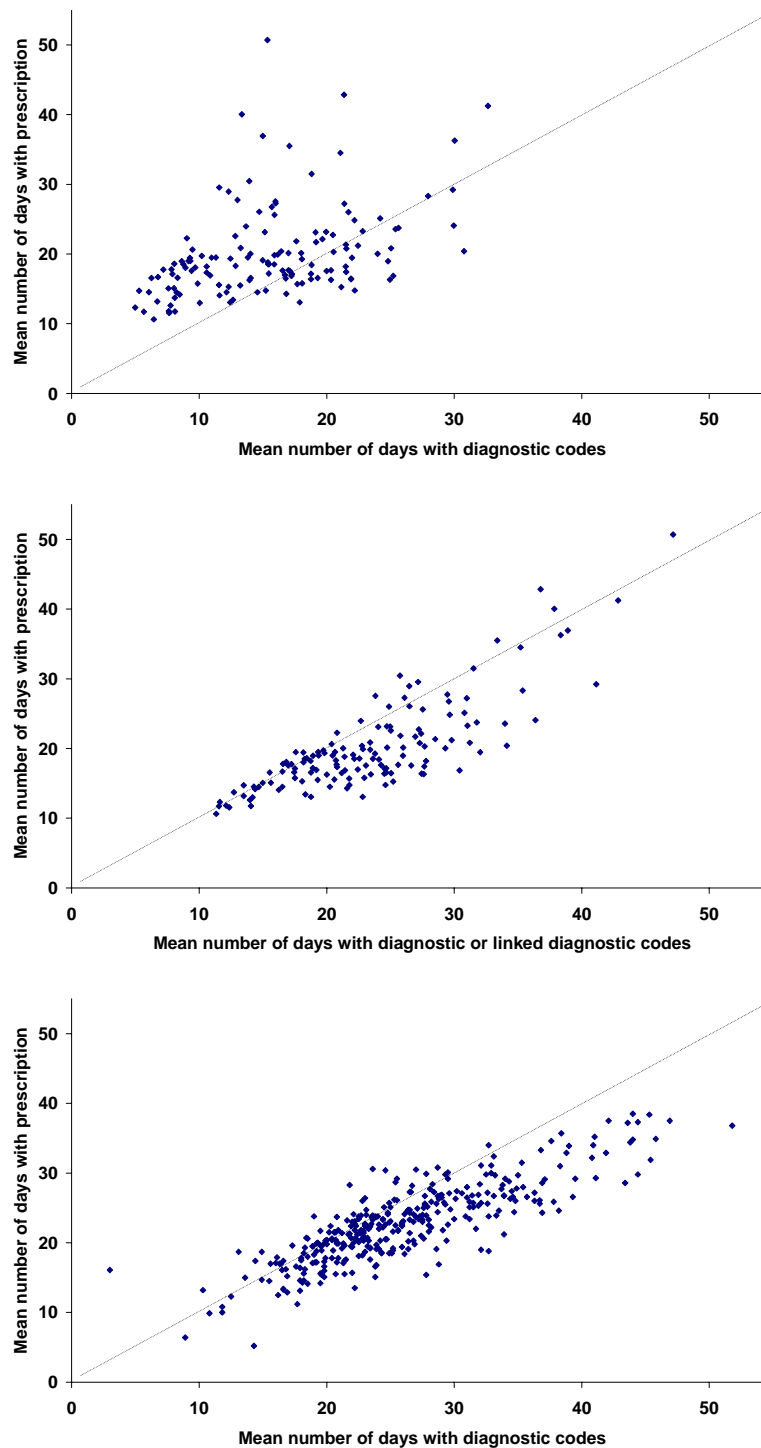


Figure 1
Scatter plots of mean number of days with a prescription versus mean number of days with a diagnosis by practice in the DIN and GPRD birth cohorts (minimum 25 children per practice) Top figure – (a) DIN: Mean number of days with a prescription issued vs. diagnostic code (without linkage, n = 140). Middle figure – (b) DIN: Mean number of days with a prescription issued vs. diagnostic or linked diagnostic code (n = 140). Bottom figure – (c) GPRD: Mean number of days with a prescription issued vs. diagnostic code (n = 437)

Table 5: Summary of emollient prescribing and eczema recording in DIN and GPRD birth cohorts in the first 5 years of life

| | DIN | | GPRD | |
|--|--------|--------------|---------|--------------|
| | Number | % | Number | % |
| Children | | | | |
| Total number of children | 40,183 | | 76,310 | |
| Of which have at least one emollient prescription | 16,799 | 41.8% | 27,955 | 36.6% |
| Children with an emollient prescription | | | | |
| Of which also have an eczema diagnosis recorded in first 5 years of life | 15,404 | 91.7% | 22,444 | 80.3% |
| Of which also have an eczema diagnosis recorded that can be linked by problem number (and thus under the same chapter heading) in first 5 years of life (DIN only) | 14,583 | 86.8% | – | |
| Emollient prescriptions days | | | | |
| Total number of days with emollient prescriptions in first 5 years of life | 85,870 | | 123,413 | |
| Of which have an eczema diagnosis that appears on the same day | 16,624 | 19.4% | 41,647 | 32.7% |
| Of which an eczema diagnosis is made on the same day or can be linked to a previous one by problem number | 74,217 | 86.4% | – | |
| Of which have a chapter heading of eczema | 64,461 | 75.1% | – | |
| Of which have a chapter heading of eczema – excluding 20 practices (see methods n= 71,701) | 59,688 | 83.3% | – | |

diagnosis and symptom codes occurring less frequently than they do in GPRD. However, this disparity is eliminated by taking account of linkage in DIN.

Comparing the databases

We know from preliminary work that the age-sex structure of DIN is comparable to GPRD for the years in which they overlap, as well as to the population of England and Wales itself. In this report, we studied the composition of the medical records in more depth by creating birth cohorts in each database. The methodology used to create these was similar in each, with the main difference being in their data collection periods. Thus DIN includes births from later years (1989–1997) compared with GPRD (1989–1993) where we only had data available to 1998. Children born into DIN in these later years had records of comparable length and composition to children born in during 1989–93, suggesting no appreciable bias was incurred by comparing the two databases outright in the analyses.

While prescribing records in the DIN and GPRD birth cohorts appeared very similar, the usage of diagnostic codes was more parsimonious in DIN. This is not unexpected, as users of the Meditel system in DIN are strongly encouraged to take a problem orientated approach to structuring the medical record. These fundamental differences explain why preliminary work we have carried out suggest that period prevalence rates of disease based on diagnostic codes alone are lower in DIN, while rates based on treated disease (requiring associated medication plus the existence of the diagnosis ever) are comparable.

In this study we attempted to account for the linkage in DIN by assuming that a diagnosis was 'present' when a prescription was made that linked into an existing diagnosis (i.e. it had the same problem heading number). This adjustment resulted in DIN producing a very similar number of 'diagnostic' days to that seen in GPRD (23.1 to 22.9).

Importantly such a definition is an indication of an ongoing medical problem, not necessarily of a consultation, as repeat prescriptions will also count as 'diagnostic' day where they link to a diagnosis. It is difficult to produce a similar definition within GPRD. However, the mean number of days on which either a prescription was issued or a diagnostic code was recorded is similar in the two databases (26 in DIN, 27.4 in GPRD).

Quality of linkage

To validate further the quality of linkage in DIN, we studied a specific example in order to highlight some of the potential problems. We chose a class of drugs (emollients) that is commonly prescribed in children, and is quite specific for the condition (eczema or dry skin) it is prescribed for.

While the majority of emollient prescriptions were linked to an eczema diagnosis this linkage was not specific. In some practices many diverse diagnoses were listed under a single problem heading. Our solution was to focus on the problem headings themselves, and to lay down more stringent requirements for what we considered 'good linkage' and a better problem oriented medical record. We

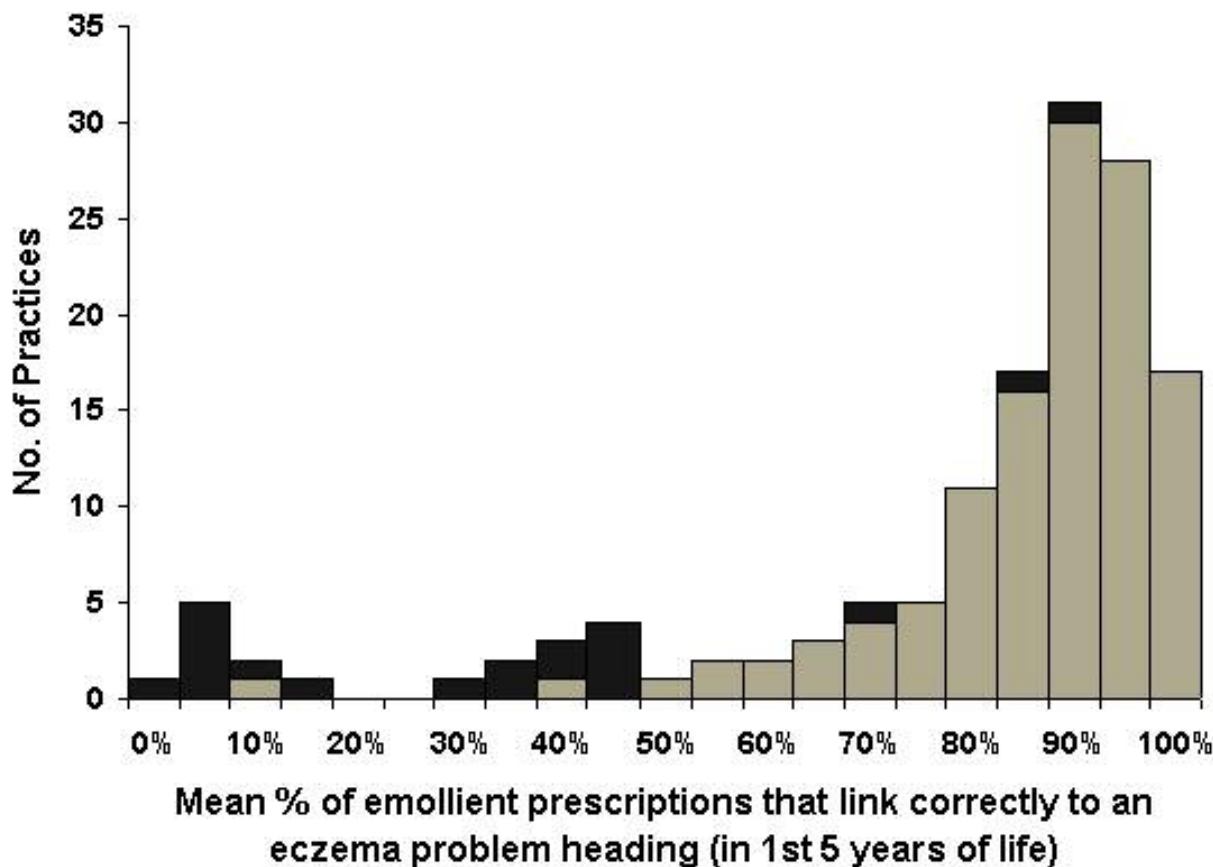


Figure 2
Practice proportions of emollient prescriptions that link to an eczema chapter heading in 1st five years of life
 Legend: Dark Shading – Not using chapter headings satisfactorily Light Shading – Using chapter headings satisfactorily

identified 20 practices that were using the system in ways that would be unsuitable for any analysis based on linkage. These practices were subsequently excluded in our final analysis, though for many purposes their data would be considered adequate.

Overall the linkage of emollient prescriptions to a satisfactory diagnosis either within the record or based on problem heading was good. However, it varied markedly by practice. Much of this variation was explained by the use of less specific levels of the Read Code hierarchy by some practices as the linkage was markedly improved by the exclusion of the 20 'unsuitable' practices. Two remaining practices, with notably sub-standard linkage of emollients to eczema, were explained by their reliance on skin-related Read Codes which did not fit our definition of eczema.

Using Read Codes which appear as the problem heading provides a single reason, and where that is a low level code, a specific reason, for the prescription. Using all diagnoses appearing under that problem heading is problematic as multiple diagnoses and symptoms may appear under the one problem heading. For emollients, we calculated that on average over five diagnostic codes were being linked to each problem number. Many of these were valid and represent the development of the condition (e.g. rash becomes eczema), however some are unrelated. A likely scenario is that when a child presents with two separate problems on the same day, the GP does not always create two separate problem entries.

The value of the Problem Orientated Medical Records for research

While the POMR was introduced as a way of improving clinical care in a secondary care setting[2,8], it only became feasible within primary care with the advent of practice computer systems able to rapidly assemble data for a patient into a number of different views[9]. The debate over whether POMR should be an integral part of the electronic patient record continues: one American committee split over whether to recommend POMR[9] while one UK group has argued for a more structured record including timelines, problems, episodes and consultations[10]. However, we are not aware of attempts to evaluate the implications for research.

Our findings emphasise that in analysing data collected using a system based on POMR it is crucial to take account of linkage if sensible period prevalence rates are to be obtained. A potential advantage of the Meditel system is that the reason for prescribing a drug should be available: this is of undoubted interest to drug companies, but also in pharmaco-epidemiology. For example, an analysis of why HRT was prescribed during the 1990's would be of considerable interest. However, differences in the way practices use problem headings - including inventing their own or using rather broad headings such as "skin problems" - raises important problems both for using linkage for research purposes (where it requires validation) and with transferring records from one GP system to another.

Whether or not POMR will be part of the electronic patient record of the future, we have a decade's worth of data collected using such a system during the 1990's. The problem orientated linkage in the DIN database offers a level of information about the relationship between diagnoses and medication that does not exist in GPRD. It also offers important advantages for understanding trends in prescribing and for feedback to GPs compared to Prescribing Analysis and Cost (PACT) data, which lack links to both demographic and clinical characteristics of those prescribed for. In theory, the POMR structure should also give insight into the evolution of diagnoses.

Conclusions

We have demonstrated the importance of comparing large-scale GP databases based on fundamentally different computing systems. While prescribing records in DIN and GPRD appeared similar, the usage of diagnostic codes is more parsimonious in DIN. However, if linkage of prescriptions to problem headings is taken account of, then the volume of diagnostic codes recorded in DIN is very similar to that in GPRD and results in similar period prevalence rates for many conditions. A potentially important advantage of the POMR structure of DIN, if used satisfac-

torily as it was in 121 of the 141 practices, is that most prescriptions can be linked directly to a diagnostic heading, providing a reason for the issue. This is but one example of the value of carrying out research in parallel in databases based on different systems. Others include the ability to validate findings found within one database and the availability of different explanatory variables (with the ability to adjust for them) such as a social indicator in DIN[4] and a family index in GPRD[11].

List of abbreviations

BNF British National Formulary

DIN Doctors Independent Network

EPR Electronic Patient Record

GPRD General Practice Research Database

IPS In Practice Systems

OXMIS Oxford Medical Information System

POMR Problem Orientated Medical Record

PPA Prescription Pricing Authority

VAMP Value Added Medical Products

Competing interests

Nicky Richards and Steve Caine are directors of Compu-File Ltd. which markets data to pharmaceutical companies.

Authors contributions

DC, DS and SH conceived the idea for this work and raised grant funding. DC, IC and SDeW developed the ideas for this paper. NR and SC advised on practical issues relating to DIN data. IC carried out the bulk of the analyses. SB carried out the GPRD analyses and helped set up the birth cohort in DIN. IC wrote the paper in collaboration with DC and SDeW. All authors commented on drafts of the paper

Acknowledgements

This work was funded by Wellcome Trust Grant 065177/Z/01/Z

References

1. Benson T: **Why general practitioners use computers and hospital doctors do not- Part I: incentives.** *BMJ* 2002, **325(7372)**:1086-1089.
2. Weed LW: **Medical Records that guide and teach.** *N Engl J Med* 1968, **278**:593-600.
3. Hollowell J: **The General Practice Research Database: quality of morbidity data.** *Popul Trends* 1997, **87**:36-40.
4. DeWilde S, Carey IM, Bremner SA, Richards N, Hilton SR and Cook DG: **Evolution of statin prescribing 1994-2001: a case of agism but not of sexism?** *Heart* 2003, **89(4)**:417-421.

5. Lawrenson R, Williams T and Farmer R: **Clinical information for research; the use of general practice databases.** *J Public Health Med* 1999, **21(3)**:299-304.
6. Office for National Statistics: **Key Health Statistics from General Practice 1998.** London: The Stationery Office 2000.
7. Bremner SA, Carey IM, DeWilde S, Richards N, Maier WC and Hilton SR et al.: **Early-life exposure to antibacterials and the subsequent development of hay fever in childhood in the UK: case-control studies using the General Practice Research Database (GPRD) and the Doctors' Independent Network (DIN).** *Clinical & Experimental Allergy* in press.
8. Weed LW: **Medical Records that guide and teach (concluded).** *N Engl J Med* 1968, **278**:652-657.
9. **The Computer Based Patient Record: an essential technology for health care.** National Academy Press 1997. Also available online at <http://www.nap.edu/books/0309055326/html/index.html>
10. Bainbridge M, Salmon P, Rappaport A, Hayes G, Williams J and Teasdale S: **The Problem Orientated Medical Record – just a little more structure to help the world go round?** . Also available online at <http://phcsg.ncl.ac.uk/conferences/camb96/mikey.htm>
11. McKeever TM, Lewis SA, Smith C, Collins J, Heatlie H and Frischer M et al.: **Siblings, multiple births, and the incidence of allergic disease: a birth cohort study using the West Midlands general practice research database.** *Thorax* 2001, **56(10)**:758-762.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2296/4/14/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

