Journal of Biomedical Informatics 43 (2010) 982–987

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

The Equity in Prescription Medicines Use Study: Using community pharmacy databases to study medicines utilisation

Simon Horsburgh [a,*], Pauline Norris [a], Gordon Becket [b], Peter Crampton [c], Bruce Arroll [d], Jacqueline Cumming [e], Peter Herbison [f], Gerald Sides [a]

[a] School of Pharmacy, University of Otago, Dunedin, New Zealand
[b] Pharmacy and Pharmaceutical Sciences, University of Central Lancashire, United Kingdom
[c] Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand
[d] Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand
[e] Health Services Research Centre, Victoria University, Wellington, New Zealand
[f] Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand

ARTICLE INFO

Article history:
Received 24 March 2010
Available online 13 August 2010

Keywords:
Drug utilisation
Medical records systems
Computerized
Medical record linkage
Pharmacoepidemiology

ABSTRACT

Purpose: Pharmacy dispensing databases provide a comprehensive source of data on medicines use free from many of the biases inherent in administrative databases. There are challenges associated with using pharmacy databases however. This paper describes the methods we used, and their performance, so that other researchers considering using pharmacy databases may benefit from our experiences.

Methods: Data were collected from all nine pharmacy dispensing databases in an isolated New Zealand town for the period October 2005–September 2006. Probabilistic record matching was used to link individuals across pharmacies. Patient addresses from the pharmacy data were geo-located to small areas so an area measure of socioeconomic deprivation could be assigned. Medicines were coded according to the ATC-DDD drug classification system.

Results: Data on 619,264 dispensings were collected. Record matching reduced an initial pool of individuals from 54,484 to 38,027. Socioeconomic deprivation ranks were assigned for 30,972 (93%) of the 33,375 unique addresses identified, or 36,048 (95%) of individuals. ATC codes were assigned to 613,490 (99%) of the dispensings, with DDDs assigned to 561,223 (91%). Overall, 93% of dispensing records had complete demographic and drug information.

Conclusions: The methods described in this paper generated a rich dataset for medicines use research. These methods, while initially resource-intensive, can to a great extent be automated and applied to other locations, and will hopefully prove useful to other researchers facing similar challenges with using pharmacy databases. However, it is difficult to envisage these methods being viable on a long-term or national scale.

1. Introduction

Medicines are a front-line medical intervention. They account for large portions of health care expenditure in OECD countries [1,2] and are a significant mediator of health status [3,4]. As such, it is important that the distribution and effects of medicines are understood so that their health benefits are both optimal and equitable, and their use affordable to modern health systems.

In many countries, researchers are reliant on medicines subsidisation data for population-level drug utilisation studies (see [5,6] for examples). Although the precise details of the various subsidisation schemes differ across countries (and even states), differential under-capture of dispensings depending on the coverage of the scheme is a common issue [7,8]. Other schemes, such as the United Kingdom's Prescription Analysis and Cost (PACT) system, do capture all dispensings but lack socio-demographic information on the patient [9]. These limitations place restrictions on the types of research which can be performed using these data. For example, research may only cover those medicines or people known to be well-captured by the scheme. Drug utilisation studies on medicines use in the elderly are a good example of this.

A similar situation exists in New Zealand. Much of the drug utilisation research performed in New Zealand has used data from the Pharmaceuticals Collection database (known as '*Pharms*'). *Pharms* is maintained by the Ministry of Health and the Pharmaceutical Management Agency (PHARMAC), and contains records of dispensings where the Government has made a contribution to the cost of (or 'subsidised') the medicine. In New Zealand, the Government

* Corresponding author. Address: School of Pharmacy, University of Otago, P.O. Box 913, Dunedin 9054, New Zealand.
E-mail address: simon.horsburgh@otago.ac.nz (S. Horsburgh).

1532-0464/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2010.08.004

pays much of the cost of most medicines. However, New Zealanders are often expected to pay a certain fixed amount towards the cost of a medicine dispensing. This amount was, until the mid-2000s, up to NZ$15 per item but is now normally NZ$3 (and sometimes nothing) and has varied according to the patient's age, income and health status. If the medicine is particularly cheap, the patient's payment may actually cover the cost of the medicine. In these cases the medicine will not be recorded in *Pharms*, as the Government's contribution was nil. Likewise, *Pharms* will not contain records for medicines where the Government does not contribute to the cost at all (as for some 'lifestyle' medicines like Xenical). Precisely what level of bias this introduces is unknown [10]. Even when patient payments are low (say NZ$3), certain common medicines may still be sufficiently inexpensive that they cost less than the patient payment. Antibiotics are a good example of this [11]. Previous research has demonstrated the uneven capture of *Pharms*, with some medicines found to be well-captured while others are not [11,12]. Longitudinal studies are also made difficult, since changes in the subsidisation status of a medicine may have a profound effect on whether it is captured by *Pharms*.

Pharmacy dispensing databases have been recognised by researchers as an alternative data source for drug utilisation studies [6,13–15]. Pharmacy dispensing databases contain records of all dispensings regardless of subsidisation status, as well as patient information. This is a significant advantage, as it removes a substantial source of bias.

There are a number of challenges with using pharmacy dispensing databases as the primary data source, however. Pharmacies in New Zealand operate independently of each other, and do not pool data (except with *Pharms*). This makes data collection difficult, as data have to be obtained from each pharmacy separately. It can also require extremely complicated sampling frames, as pharmacies can vary considerably in their client base. Patients often use multiple pharmacies, which means that the records for a patient in any one pharmacy may be incomplete [13]. Pharmacy dispensing databases do not generally include information diagnosis to dispensing, and do not reliably include non-prescription medicines sold 'over-the-counter'.

This is a common scenario internationally. There are a limited number of places with systems for collecting comprehensive dispensing data from all pharmacies in an area. These systems, when they are available to drug utilisation researchers, have been tremendously fruitful (see [5,16,14,15]). For the majority of researchers, including ourselves, who work in areas without such systems, other approaches need to be developed.

The Equity in Prescription Medicines Use (EIPMU) study was undertaken to examine medicines use among groups in the population in order to identify whether such use matched the health needs of those groups. Since the probability of a dispensing being subsidised was associated with being in one of these population groups, pharmacy dispensing databases were used instead of *Pharms*. We sought to overcome some of the challenges of using pharmacy dispensing databases by collecting data from all of the pharmacies in a medium-sized town. Since the town was a substantial distance from any other town with a pharmacy, it was hoped that this would provide almost complete capture of medicines dispensings to people from that town and thereby avoid the possible biases associated with sampling pharmacies. The EIPMU built upon our experiences of using similar methods in a smaller, earlier study [13]. However, the process was not straight-forward.

This paper will describe the methods used in the EIPMU study. Results from the study will be presented in future papers. We present the methods separately here to share our approach and techniques with other researchers considering using unlinked pharmacy databases in their drug utilisation research.

2. Methods

The EIPMU study took place in Gisborne, a town servicing the Gisborne Region in the north east of New Zealand. The Gisborne Region has a population of 44,463 people, of which 41,922 (94%) live in Gisborne city [17]. This region was chosen as the study centre because of the following:

- It was geographically isolated, with the nearest pharmacy outside the district being more than an hour's drive away from any other pharmacy in the Region. We could therefore be confident that almost all of the dispensings to the Region's inhabitants were provided by the Region's pharmacies.
- The Gisborne Region had a high proportion of Māori (44%), the indigenous people of New Zealand. It also had a mix of people across the spectrum of socioeconomic positions. This improved our ability to examine equity issues by ethnicity and socioeconomic deprivation.

The study covered the year 1 October 2005 to 30 September 2006 inclusive. Ethical approval for the study was granted by a Ministry of Health accredited ethics committee.

2.1. Data collection

Data were collected from all eight of Gisborne's community pharmacies. Records of outpatient dispensings from the hospital pharmacy were also obtained. Community pharmacies all used either the L.O.T.S [18] or Toniq [19] dispensing systems, while the hospital pharmacy used a bespoke-system. The data from community pharmacies were collected by physically accessing the computer system and downloading the data from the pharmacy's dispensing software. The data collected included (but were not limited to):

- Patient demographics, including name, date of birth, gender, address and health system identifier (National Health Index, described below).
- Details about the dispensed medicine, such as the medicine, amount dispensed and medicine strength.
- Financial information about the dispensing, including the cost to the patient, the amount paid by the Government, any fees associated with the dispensing and any healthcare concessions applied.

A patient's first initial, surname and address must be recorded for every dispensing by a pharmacy under the Medicines Regulations (1984). Date of birth is also required for patients under the age of 13. This is regardless of whether the medicine being dispensed is subsidised or not. Emergency contraception dispensings are an exception to this, where the patient can request that personal information not be recorded.

The data files were encrypted immediately at the pharmacy to ensure the confidentiality of the data. Data from the hospital pharmacy were provided electronically by their staff, and provided less financial information than the data from the community pharmacies.

2.2. Record matching

Previous research has shown that people often patronise more than one pharmacy [13]. This meant that individuals had to be identified across pharmacy datasets. New Zealand has a health system identifier (the National Health Index, or NHI) which is assigned to a person at first contact with the health system

(generally birth). However, we discovered that the NHI by itself would not be adequate to match people across pharmacies because:

- The NHI is supposedly unique to an individual and each person should only have one. However, some people have more than one and different NHIs are sometimes recorded in different pharmacies. Checking of NHIs revealed that 7% of NHIs (10% of records) recorded were instances where the same person had multiple NHIs.
- Pharmacists in New Zealand are not required to record the health system identifier, although most do. Eighty-eight percent of the records from the study pharmacies had an NHI recorded.
- A preliminary check revealed that 2% of the NHIs (4% of records) recorded did not pass the check algorithm for determining whether an NHI is of a valid form, or was used for several different patients.

Taken together, the above indicated that matching on NHI alone would be unwise. We therefore decided to use a probabilistic matching process, combining the patient information from all of the pharmacies and treating it as a de-duplication problem. It is not uncommon for people of non-European ethnicities, particularly Māori and Pacific people, to use both their given names and an anglicised version. We therefore decided to manually review all pairs to gain some insight into any issues this might cause with de-duplication. The steps of this process are outlined below.

Step 1 – Initial probabilistic de-duplication

A dataset of unique combinations of NHI, name, date of birth and gender was created. This initial dataset was then de-duplicated using the above variables with the freely-available software package LinkPlus [20]. LinkPlus employs a probabilistic Expectation Maximisation (EM) algorithm to assign weightings based on the similarity between two records, accounting for name frequency in the dataset [21].

The matched pairs of records output by LinkPlus were manually reviewed by two study team members. Definite matches and non-matches were marked as such. Uncertain matches were treated as non-matches. Matched records belonging to the same person were assigned a computer-generated unique identifier. Gender and dates of birth which had been missing from a person record were filled in if that information was available from another record which had been definitely matched.

Step 2 – Initial linkage to the NHI dataset

Once the initial de-duplication step had completed, the initial unique person combinations were sent to the Information Directorate of the Ministry of Health[1] (ID) for linking to the NHI database. The ID is a Government agency which maintains many of the health information databases in New Zealand. This includes the NHI database, which contains a list of NHIs and demographic information about the people they belong to. This demographic information includes the person's name, date of birth, gender, ethnicity, socioeconomic deprivation and address.

Records from the EIPMU dataset were matched to the NHI database by ID staff using a direct matching algorithm based on name, date of birth, gender and NHI (if available). The records which were successfully matched were then returned augmented with name, date of birth, gender and NHI data from the NHI. This additional information from the NHI was then used to fill in missing data from the pharmacy records and disambiguate inconsistent data (such as when two records obviously referring to the same person had different dates of birth). The filling in of missing data and correcting inconsistent records was made possible through the use of the unique person identifiers assigned in Step 1.

Step 3 – Second probabilistic de-duplication

The de-duplication process of Step 1 was repeated using the augmented person records from Step 2. Matched records belonging to the same person were assigned a new unique identifier. Re-running the de-duplication process on the augmented records was performed to maximise de-duplication yield by increasing the amount and quality of the information available to the de-duplication algorithm.

Step 4 – Second linkage to the NHI dataset

The records from Step 3 were sent to the ID, where the ID again matched the EIPMU records to those in the NHI database using the methods described in Step 2. The records were returned augmented with name, date of birth, gender, NHI, ethnicity, address and date of death from the NHI database. The primary purpose of this Step was not to improve the de-duplication quality, but to obtain patient information such as ethnicity, which is not recorded in pharmacy dispensing databases.

2.3. Assigning socioeconomic deprivation ranks

A key aim of the EIPMU study was to investigate how medicines use varied by socioeconomic deprivation. We used the NZ Deprivation Index (NZDep2006, the 2006 version) as our measure of socioeconomic deprivation [22,23]. NZDep2006 is a small area measure of material socioeconomic deprivation which is widely-used in New Zealand and has been validated in previous studies [23–25]. The NZDep2006 assigns a deprivation ranking to a small area based on certain demographics of the area residents. In order to assign a NZDep2006 ranking to an individual it is necessary to ascertain which small area they reside in. Unlike countries such as the United Kingdom, New Zealand does not have post-codes which precisely locate an address. The individual's whole address therefore needs to be parsed and placed within the correct small area.

Community pharmacies in New Zealand routinely record patient addresses. The addresses are recorded for use by the people working in the pharmacy, however, and not for automated processing by computers. The formatting of addresses therefore varies tremendously between, and often within, pharmacies. This presented the challenge of geo-locating these addresses to the correct small area. The process we followed is outlined below.

Step 1 – Address standardisation

The address for each person was converted into a standard form which could then be matched to an area. This process was performed using the address standardisation routines of the Freely Available Biomedical Record Linkage programme (FEBRL) version 0.4 [26]. FEBRL uses Hidden Markov Models (HMM) to 'learn' patterns of address formatting [27]. This allows the HMM to be trained with a subset of addresses and then used to output standardised addresses when applied to the complete set. HMMs are able to generalise to novel input based on the rules learned from the training set.

An HMM was trained for each pharmacy, as there was substantial variation in the formatting of addresses between pharmacies. Each HMM was trained with an initial set of 250 unique addresses from that pharmacy. When these addresses were being output to the correct format, the HMM was applied to the complete list of unique addresses from the pharmacy. The quality of the output formatting was checked and the first few records which were not correctly formatted were added to the training set. The lookup tables for street and institution names were also updated if needed. This process was repeated until the accuracy of the HMM could no longer be improved. Typically this process led to output which was all in the correct format with a training set of between 300 and 600 records. At worst, the HMM was unable to correctly standardise one or two addresses.

[1] Formerly the New Zealand Health Information Service, or NZHIS.

Step 2 – Matching to areas

Once the addresses had been standardised, they were matched electronically to a small area through a concordance list which had been obtained from Statistics NZ [28]. Not all addresses gave sufficient detail to automatically match through the concordance list, and some addresses could apply to several places. For example, institutions and farm names were often recorded, and the study area had three geographically-distinct Main Streets. The addresses for such records were manually determined if possible through searches of the web (Google Maps, Google, the online Regional Council rates database) and telephone books for the area.

Step 3 – Assignment of deprivation ranks

Once an address had been assigned to a specific area, the person was assigned the deprivation rank for that area. If the person had several different addresses, the mean of the deprivation ranks for each unique address was assigned.

2.4. Medicines coding

Pharmacy dispensing databases in New Zealand contain information on the name (brand and generic), strength and quantity of the medicine dispensed. They also contain a pharmacode, which is a unique numeric identifier for a specific drug product. Pharmacodes are administered by the New Zealand Pharmacy Guild and every licensed medicine in New Zealand is assigned one.

The databases did not contain an Anatomical Therapeutic Chemical (ATC) code for each medicine. The ATC [29] is an international standard for classifying medicines, and is used extensively in drug utilisation research. Several colleagues had assigned an ATC code to all licensed medicines in New Zealand [30]. Using this information, we were able to create a conversion table so that pharmacy database pharmacodes could be automatically converted into ATC codes.

The total quantity of medicine dispensed in each dispensing was calculated in Defined Daily Doses (DDDs) [29] for all medicines which were assigned an ATC code. The DDD system provides a means for scaling quantities of different medicines into broadly equivalent units whilst accounting for the relative potencies of the active ingredients. Each record had the total number of DDDs contributed by that dispensing calculated when possible.

Compounded medicines (that is, medicines made or mixed by the pharmacist) were not ATC or DDD coded.

3. Results

3.1. Data collection

A total of 654,129 dispensing records were obtained. Dispensings which were not to individuals (such as those for doctor's supplies or to institutions), as well as those for animals, service fees or over-the-counter sales were removed from the dataset. This left 625,288 (96%) dispensing records. After record matching, a further 6024 (1%) records were removed because the individuals did not reside in the study area. This left 619,264 (95% of original total) dispensing records for analysis.

Unless otherwise stated, individuals residing outside the study area are excluded from the figures presented in these analyses.

3.2. Record matching

Initially, 54,484 individuals were identified through unique combinations of NHI, name, date of birth and gender. This number was reduced in Step 1 to 39,100 individuals, a reduction of 28%. After receiving additional information through linkage to the NHI database, 38,027 individuals were identified in Step 3. This was a 3% improvement.

Matching to the NHI database in Step 2 was highly successful, with 44,395 (81%) of the records sent matched. After the integration of the de-duplication process results from Step 3, 35,455 (93%) of the records sent were matched to a NHI record.

Table 1 shows the effect of the record matching process on data coverage in dispensing records for age, gender and NHI. Age and NHI coverage were already high in the raw pharmacy data. Record matching only led to modest improvements, although it should be noted that the NHIs in the post record matching figure had been confirmed to match a record in the NHI dataset, whereas those in the raw pharmacy data had not. Gender was less well-recorded in the raw pharmacy data, but record matching dramatically improved the coverage achieved.

3.3. Assigning socioeconomic deprivation ranks

Of the 33,375 unique addresses recorded, 1386 (4%) were adjudged impossible to assign to an area due to lack of information (e.g. 'Gisborne') or being ambiguous (e.g. '123 Main Road', where there are several Main Roads in the Gisborne region with that street number). A total of 30,972 (93%) were able to be matched to an area. This led to deprivation ranks being assigned for 36,048 (95%) individuals after record matching had taken place.

3.4. Medicines coding

ATC codes were assigned to 613,490 (99%) of the dispensing records. Of the 5774 (1%) not assigned, 2579 (45%) were compounded items (i.e. items such as creams made up by the pharmacist). The DDD was able to be calculated for 561,223 (91%) dispensing records. Of the 58,041 (9%) not assigned, 52,186 (90%) did not have a DDD for that medicine in the ATC classification, and 5774 (10%) could not be assigned because no ATC was able to be assigned to the medicine. The remaining 81 (0%) dispensing records could not be assigned a DDD because there was insufficient information on formulation strength.

3.5. Data coverage

The ultimate data coverage for the key variables of interest is summarised in Table 2. Coverage for demographic variables was very high, with 88% of individuals and 93% of dispensing records containing information on all four demographic variables. Some clustering was evident when information was missing for more than one variable. Missing data for both the ethnicity and age variables was the most common combination, with 37% of individuals with missing data having information for both those variables missing. The next most common combination was ethnicity and gender (24%), followed by age and gender (15%). Overall, two percent of the individuals with missing demographic information had missing data for only one variable.

ATC coverage was extremely high, with very good total DDD coverage. The final figure for dispensing records with complete person and drug information understates the actual success of this process. Not all ATC categories have a DDD to assign. When re-

Table 1
Improvement in age, gender and NHI coverage resulting from record matching.

Demographic	Raw data		Post record matching		Improvement (%)
	n	%	*n*	%	
Age	575,130	93	614,069	99	7
Gender	493,844	80	608,126	98	23
NHI	559,946	90	603,502	97	8

Table 2
Summary of demographic and medicine data coverage

	Dispensings		Individuals	
	n	%	*n*	%
Total	619,264		38,027	
Demographic				
Age	614,069	99	36,283	95
Gender	608,126	98	36,907	97
Ethnicity	598,305	97	34,931	92
Socioeconomic deprivation	595,348	96	36,048	95
All present	576,661	93	33,275	88
Medicine				
ATC	613,490	99		
Total DDD	561,223	91		
Complete	522,563	84		

calculated using only the ATC categories where a DDD was assignable, 93% of dispensing records had complete person and drug information.

4. Discussion

This paper has described the methods we used to obtain comprehensive data on medicines dispensings directly from pharmacies. In particular, it has focussed on three key challenges that we encountered in using this approach: matching individuals across the different pharmacies, obtaining an area deprivation measure from the raw addresses recorded in the pharmacy databases, and coding the medicines information into standard international classifications. The methods used have resulted in a rich dataset with which to examine a wide range of drug utilisation and access research questions. Data coverage was very high for patient demographics, and good for standardised medicines classification. With the improvement in NHI coverage as a result of the record matching process, the dataset also holds potential for future pharmacoepidemiological research through linkage to other health datasets (such as hospitalisations and accident insurance claims).

A key limitation of the methods described here was that they were very time-consuming. The collection of the data was the least intensive part, and could easily be performed by pharmacy staff. The subsequent data processing and cleaning accounted for the majority of the effort due to the large amount of manual review and intervention required because of inconsistencies and errors in patient information. This is to be expected; dispensing databases are used to run a pharmacy, not support the needs of researchers [31]. Inconsistencies in this information are of little consequence in the daily operation of a pharmacy, because staff are easily able to deal with them. Information on the medicines themselves and their cost tended to be more consistently recorded, as these are automatically generated by the dispensing software itself and are required for reimbursement of subsidised medicines costs.

There is an argument to be made for the standardisation of patient information elements and their recording in pharmacy databases. However, enforcing such standardisation in dispensing software comes at the cost of hindering pharmacist workflow. Anecdotal evidence from dispensing software manufacturers suggests that pharmacists often turn off features designed to improve data entry (such as checks to ensure that an entered NHI is in a valid format) because they are seen as annoying interruptions. At a minimum, improving NHI coverage and quality would significantly reduce the effort needed to use pharmacy databases for research.

There have been some moves in New Zealand towards improving the coverage and quality of NHI recording in pharmacy dispensing databases as part of eHealth initiatives [10], given the role of the NHI in linking information about a patient across the health system. There is a requirement to have NHIs recorded on 90% of the dispensings submitted for Government reimbursement (Kathy Blake, Ministry of Health Sector Services, personal communication, 3 February 2010), but pharmacists are not required to provide the NHI if it is not recorded and there is no checking of the accuracy of these NHIs by the Ministry of Health beyond ensuring that they are in the correct format. An incorrectly entered NHI with the correct format, or one referring to the wrong patient, will not be detected. There is therefore little real incentive for pharmacies to improve their recording of NHIs, and little feedback available for them to correct errors.

An important goal of this study has been to automate the methods we developed for processing and cleaning the pharmacy dispensing data. This has to a large extent been achieved. The coding of medicines from dispensing databases has been automated, as has much of the record cleaning. The trained HMMs are re-usable for the Gisborne area should the study be repeated there, and should generalise with minimal re-training to other areas. The main point in the process still requiring a reasonable time investment is the matching of patient records across pharmacies. This has the potential to be automated by selecting threshold scores for the automatic assigning of match and non-match status, and then undertaking a manual review of the subset of records between these thresholds. Such an approach was not used in this study, as we wanted to create a dataset as close as possible to being correctly de-duplicated. Improvements in NHI coverage and accuracy would make this process substantially more efficient.

The challenges faced in generating this dataset are not unique, and in many ways reflect those encountered in the initial stages of other now well-established automated pharmacy dispensing data systems [16,15]. The availability (or development) of a reliable personal healthcare identifier in pharmacy databases was a critical part of making these systems viable on an on-going basis. While the processes used in this study could be repeated reasonably efficiently, it is difficult to see them as being viable on a long-term (or ultimately national) scale with the current coverage and quality of NHIs in the pharmacy databases. The consistent and accurate recording of a personal healthcare identifier in pharmacy dispensing databases are key requirements before a viable, long-term automated pharmacy dispensing data system can be developed and deployed. In the New Zealand context, this means that pharmacies need to be required to record accurate NHIs in their dispensing databases, or incentivised to do so.

Summary table

What was already known on the topic:

- Pharmacy dispensing databases can provide a rich source of medicines use and pharmacoepidemiological data free of many of the biases inherent in administrative databases.
- Pharmacy dispensing data are often not collated, making it difficult to obtain comprehensive capture of drug exposure.
- Pharmacy dispensing databases are primarily intended for day-to-day use in a pharmacy. It can take a substantial amount of processing to make the data usable for research purposes.

What this study added to our knowledge:

- This study has described methods for collecting and processing key information from pharmacy dispensing databases which were successful in New Zealand.
- These methods should be broadly transferrable to other countries where comprehensive medicines use data are needed, but data from pharmacy databases are not shared and centrally accessible.

Statement on conflicts of interest

There are no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the assistance of Simon Ross from the Information Directorate of the Ministry of Health.

The Equity in Prescription Medicines Use Study is funded by a Grant (07/139) from the Health Research Council of New Zealand.

References

[1] Clemente J, Marcuello C, Montañés A. Pharmaceutical expenditure, total health-care expenditure and GDP. Health Econ 2008;17:1187–206.
[2] Okunade AA, Suraratdecha C. The pervasiveness of pharmaceutical expenditure inertia in the OECD countries. Social Sci Med 2006;63:225–38.
[3] Caliskan Z. The relationship between pharmaceutical expenditure and life expectancy: evidence from 21 OECD – countries. Appl Econ Lett 2009;16:1651.
[4] Miller RDJ, Frech HEI. Is there a link between pharmaceutical consumption and improved health in OECD countries? PharmacoEconomics 2000;18:33–45.
[5] Stergachis A, Saunders KW, Davis RL, Kimmel SE, Schinnar R, Chan KA, et al. Examples of automated databases. In: Textbook of pharmacoepidemiology. Chichester: John Wiley & Sons; 2006. p. 173–214.
[6] WHO International Working Group for Drug Statistics Methodology, WHO Collaborating Centre for Drug Statistics Methodology. WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacological Services. Introduction to drug utilization research. Geneva, World Health Organization; 2003.
[7] Boudreau DM, Doescher MP, Saver BG, Jackson JE, Fishman PA. Reliability of Group Health Cooperative automated pharmacy data by drug benefit status. Pharmacoepidem. Drug Safe. 2005;14:877–84.
[8] Strom BL. Overview of automated databases in pharmacoepidemiology. In: Textbook of pharmacoepidemiology. Chichester: John Wiley & Sons; 2006. p. 167–71.
[9] Chapman SR. Prescribing information systems; making sense of primary care data. J Clin Pharm Ther 2001;26:235–9.
[10] Horsburgh S, Malik M, Norris P, Harrison-Woolrych M, Tordoff J, Becket G, et al. Prescribing and dispensing data sources in New Zealand: their usage and future directions. Dunedin, New Zealand: School of Pharmacy, University of Otago; 2009.
[11] Norris P, Funke S, Becket G, Ecke D, Reiter L, Herbison P. How many antibiotic prescriptions are unsubsidised in New Zealand? N Z Med J 2006;119:U1951.
[12] MidCentral District Health Board. Horizons medicines utilisation project. Palmerston North, New Zealand, MidCentral District Health Board; 2004.
[13] Ryan K, Norris P, Becket G. Capturing data on medicines usage: the potential of community pharmacy databases. N Z Med J 2005;118:U1677.
[14] Monster TBM, Janssen WMT, de Jong PE, de Jong-van den Berg LTW, for the PREVEND Study Group. Pharmacy data in epidemiological studies: an easy to obtain and reliable tool. Pharmacoepidem Drug Safe 2002;11:379–84.
[15] Tobi H, van den Berg P, de Jong-van den Berg L. The InterAction Database: synergy of science and practice in pharmacy. In: Medical data analysis. Berlin: Springer; 2000. p. 93–108.
[16] Evans JMM, McNaughton D, Donnan PT, MacDonald TM. Pharmacoepidemiological research at the Medicines Monitoring Unit Scotland: data protection and confidentiality. Pharmacoepidemiol Drug Saf 2001;10:669–73.
[17] Statistics New Zealand. QuickStats about Gisborne District. Statistics New Zealand; 2009.
[18] HealthSoft. L.O.T.S Dispensary; 2010.
[19] Toniq Limited. Toniq dispensary; 2010.
[20] Centers for Disease Control and Prevention. Registry Plus LinkPlus, Atlanta, Centers for Disease Control and Prevention; 2007.
[21] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc 1977;39:1–38.
[22] Salmond C, Crampton P, Atkinson J. NZDep2006 index of deprivation, Wellington. Department of Public Health, University of Otago Wellington; 2007.
[23] Salmond C, Crampton P, Sutton F. NZDep91: a new index of deprivation. Aust N Z J Public Health 1998;22:95–7.
[24] Ministry of Health. A portrait of health: key results of the 2002/03 New Zealand Health Survey. Wellington, Ministry of Health; 2004.
[25] Salmond C, Crampton P. Heterogeneity of deprivation within very small areas. J Epidemiol Community Health 2002;56:669–70.
[26] Christen P. Febrl – a freely available record linkage system with a graphical user interface. In: Proceedings of the Australasian workshop on health data and knowledge management (HDKM), Wollongong, Australia; 2008.
[27] Churches T, Christen P, Lim K, Zhu J. Preparation of name and address data for record linkage using hidden Markov models. BMC Med Inform Decis Mak 2002;2:9.
[28] Statistics New Zealand. StreetLink – Statistics New Zealand; 2009.
[29] WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment 2007, Oslo, WHO Collaborating Centre for Drug Statistics Methodology; 2006.
[30] Aaltonen K, Ragupathy R, Tordoff J, Reith D, Norris P. The impact of pharmaceutical cost containment policies on the range of medicines available and subsidized in Finland and New Zealand. Value Health 2010;13:148–56.
[31] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.