Statistical Linkage of Treatment to Diagnosis for Research and Monitoring of Practice Patterns

E. Falkoe¹, K. B. Rasmussen², M. Maclure³, H. Schroll¹
¹Research Unit of General Practice, University of Southern Denmark, Odense, Denmark
²Department of Organization and Management, University of Southern Denmark, Odense, Denmark
³Epidemiology Department, Harvard School of Public Health, Boston, MA, USA

Summary

Objectives: At each patient contact general practitioners enter information about the diagnoses and the interventions in the electronic medical record (EMR) system. If there is only one diagnosis during a single patient-physician contact, then a causal connection between the diagnosis and the intervention is established. Otherwise it is uncertain what may have been the cause of the intervention.

Methods: Ideally the general practitioners would record a match between each intervention and the diagnosis that justifies it. However, most EMR software is not capable of recording explanatory matches. Furthermore, supplying the matches is a resource-demanding task. In this study the general practitioners were supplied with a matching module for the EMR, so data with full matching between intervention and diagnosis was collected (our “gold standard”). The study models how close the full matching can be recreated by model linkage using different kinds of simple assumptions.

Results: The modeling demonstrates a raise in the measure of prediction (sensitivity) from 41.9 percent for a completely random linkage to 90.9 percent based on simple assumptions in the model. The small substantial potential further gain makes it less attractive to apply more intricate assumptions or use more complex modeling (including neural networks).

Conclusions: The perspective of the study lies in the support of the general practitioners with software for comparison of their decisions with those of their peers and also with guidelines. Thus a system for simple quality assurance and awareness to untypical decisions could be incorporated into the electronic medical record systems.

Keywords
General practitioner, family practice, classification, quality assurance, linkage, EMR data

Objectives

To use electronic medical record (EMR) data for studying how general practitioners (GPs) handle patient problems, a key step is to match each intervention to a problem or diagnosis that justifies it.

If the GP records only one problem or diagnosis and only one intervention during a single patient-physician contact, then a causal connection between the problem and the intervention is highly likely. However, when the record shows two or more problems, it is uncertain what may have been the cause of the intervention. Ideally GPs would record a match between each intervention and the problem or diagnosis that justifies it. However, most EMR software does not enable such explanatory connections to be recorded easily and quickly. This study investigates if such links can be recreated after the data have been centralized [1]. In parallel with an earlier project on methods for registration and analyses of episodes of care in general practice [2] this study classifies elderly patients’ contacts to their General Practitioner using codes for diagnoses and prescriptions. The focus of the present study [3] is on the assessment of the feasibility of using statistical methods for retrospective explanatory linkage of prescriptions to diagnoses.

Methods

Setting

The study included 42 GPs in 24 general practices in Denmark who agreed to complete an electronic form called a “matching module” for all patients aged 70 years and older, who were served between Dec. 24, 1996 and Dec. 24, 1997. Like 89 percent of general practices in Denmark today [4], these practices were fully computerized. Routinely every patient-physician contact is recorded in the patient’s EMR. Contacts include patient visits to the physician’s office, telephone consultations, and physician visits to the patient at home.

Data

For each contact the GPs routinely entered diagnoses and prescription drugs into the EMRs by selecting from pop-up alphabetic lists of diagnoses and drugs. The software required at least one diagnosis to be recorded for every contact. The EMR software recorded diagnoses addressed by the Extended Danish International Classification of Primary Care [5, 6] and interventions as medical prescriptions coded by Anatomical Therapeutic Chemical (ATC) Classification System [7]. These data were automatically downloaded with the patient number and contact date to the matching module. Later that day, the GPs directly entered the matches between prescriptions and diagnoses, indicating their reasons for prescribing the drugs.

Data were extracted from the linkage module in each general practice for all patients in the cohort, yielding 25,707 patient-physician contacts. In 15,217 contacts, there was at least one prescription recorded (26,015 prescriptions in total). In 10,941 (71.9 percent) of these contacts, there was only one diagnosis (with 11,769 prescriptions in total). In the other 4,276 contacts,
there were 9,973 diagnoses (2.3 diagnoses per contact) and 14,246 prescriptions. 2,624 diagnoses of these were not connected with a prescription. This extracted project database was our “gold standard” for evaluating the success of our statistical linkage because every prescription had a separate record containing one diagnosis that justified it. The data were structured hierarchically: each contact could have several diagnoses, and each diagnosis could have several prescriptions, but each prescription was required to have one and only one diagnosis. There exist contacts where a prescription can be said to be rationalized by more than one diagnosis (example: beta-receptor-blocking drugs for hypertension and migraine) however this is fairly seldom and we accept the simplification of our model.

Central administrative databases used by researchers to study how GPs handle the problems of contacts normally do not have diagnoses linked to prescriptions. The main variables available for linking cause and effect are patient and date of contact. To simulate this, we used the extracted database to create an unlinked combination database comprising all possible pairs (48,648) of prescriptions and diagnoses that occurred during the same contacts. For example, a contact with two prescriptions and three diagnoses contributed 6 combination pairs.

Borrowing from terminology of screening [10], we use the following definitions:

d) a true link is a match, that is linked,
e) a false link is a non-match that we mistakenly counted as a link,
f) a false non-link is a pair that we did not count as a link when in fact it is a match,
g) a true non-link is non-match, that is not linked,
h) sensitivity is the proportion of matches that we correctly linked,
i) specificity is the proportion of non-matches that we correctly do not link,
j) positive predictive value is the proportion of links that are true links, and
k) negative predictive value is the proportion of non-links that are true non-links.

**Linkage Algorithms**

Whereas record linkage normally involves comparisons of the same variables (e.g. name, sex, age) from different files, our explanatory linkage involves comparison of different variables (diagnosis and prescription) from the same file. Therefore, we hypothesized there is a need to develop algorithms, and possibly theory, for explanatory linkage. We compared the performance of the following methods of linkage from simplest to most complex.

1. **Linkage by Contact Information Only**

   As stated above, using just contact information (patient, clinic, date) we can identify about half the comparisons in the linkage database as matches because they have only one diagnosis which is almost certainly the explanation for any prescription during that contact. When a contact involves two diagnoses (D, E) and one prescribed medicine (M), then a main diagnosis is to be chosen. There are two possible links between diagnosis and medicine (DM versus EM). Without additional information, the two diagnoses appear to a computer as equally likely as the main explanation for the prescription. We say these diagnoses are tied with equal probability (here 50 percent) of being the main explanation. In algorithm 1, we consider all diagnoses within a single patient-contact as equivalent potential explanations for each prescription, and instruct the computer to choose one diagnosis for each medication at random.

2. **Linkage by Prescribing Probability in Single-Diagnosis Contacts**

   In this algorithm, we use records of contacts with only one diagnosis and one or more prescriptions to calculate the proportion of these contacts with diagnosis, D, that have a prescription for medicine, M. We use this proportion as an estimate of the probability of M given D, written Pr(M|D), in contacts with more than one diagnosis.

   For example, Table 1 shows an example of a specific contact having four prescriptions and four diagnoses. The diagnoses are: Neck syndrome, Heart failure, Diabetes mellitus and Sleep disturbance. The first diagnosis is not matched to any of the four prescriptions. However the table shows that hypnotics are in other contacts sometimes prescribed for the diagnosis of neck syndrome. The probability, given diabetes, of receiving sulfonylurea is 0.5968; this is calculated as a proportion from the combination database using all pairs from single-diagnosis contacts with the diabetes diagnosis. When no number is present, it signifies missing data due to absence of that diagnosis-prescription combination in the database among the single-diagnosis contacts.

   For each prescription in each contact, the probabilities of all diagnosis-prescription pairs are compared, and the diagnosis with the highest probability is linked to the prescription, i.e. chosen as the main explanation for the prescription. If there is a tie, i.e. the highest probabilities are equal or all probabilities are missing, the selection is made randomly. The example in the table shows that in this specific contact all the matches made by the GP are correctly linked by the algorithm.

3. **Linkage by Diagnosis Probability in Single-Prescription Contacts**

   This algorithm was identical to algorithm 2 above except we used the opposite prob-
ability, \( Pr(D|M) \), the probability that the patient had a particular diagnosis, given a particular medicine. For example, given sulfonylurea, the probability of diabetes is 100 percent. As in algorithm 2 we calculated the proportion of \( D \) among \( M \) in the contacts with only one diagnosis and one or more prescriptions.

### 4. Linkage by Combining Both Probabilities

Probabilities \( (Pr) \) combine by multiplying their complements \((1 – Pr)\). Accordingly, we produced the score \( 1 – (1 – Pr(M|D))(1 – Pr(D|M)) \). In the above example of diabetes and sulfonylurea this equals \( 1 – 0.5968 \times 0.0 = 1 \). In this algorithm, we applied this joint probability to all pairs in the linkage database and, as above, chose the pairs with the highest scores to be links.

### 5. Linkage Using All Records

Algorithms 2, 3, and 4 all estimate probabilities using data from patient-contacts with only one diagnosis and one or more prescriptions. This gives zero weight to data from patient-contacts with more than one diagnosis, which ignores potentially valuable information. On the other hand this method ensures that we are not using the same information we want to predict. Several more complicated algorithms that used the entire linkage database were tried. Some algorithms weighted every type of contact equally, while others gave proportionately less weight to the more ambiguous contacts with multiple diagnoses and prescriptions.

### Assessing Performance of Algorithms

After classifying all the pairs in the combination database as links or non-links, the results of alternative algorithms were compared by using the “gold standard” with information about the true matches. The results were tabulated as shown in Table 2 (example obtained from application of algorithm 3 based on “diagnosis probability in single-prescription contacts”, see Table 3). The format of Table 2 is identical to standard tables for validity of screening tests, from which the sensitivity and specificity, and the positive and negative predictive values, are calculated.

The sensitivity is calculated as the correctly linked matches (12,950) of the total number of matches (14,246) which results in the 90.9 percent. Note the symmetry about the diagonal in Table 2. This differs from screening test validity data. The symmetry arises from the requirement that the number of links and the number of matches both equal the total number of prescriptions, 14,246. This forces the sensitivity to equal the positive predictive value, and the specificity to equal the negative predictive value. Furthermore there is a relationship between sensitivity and specificity. Confidence intervals for the sensitivities were roughly estimated by assuming that the classification of matches as links was a series of independent binomial processes. This assumption was valid for algorithm 1 because it used random numbers to allocate medicines to diagnoses. The greatest variance will occur for the 0.5 probability and when this is applied for the completely random algorithm 1 the 95% confidence interval for the number of true-matches is from 5,847 to 6,080. Even though the other algorithms include some interdependencies that would make the true confidence intervals slightly larger than those calculated, we can conclude that the confidence interval around the figure 12,950 will not overlap the interval for the random algorithm. Our simple model thus significantly improves our prediction.
Results

The main findings are shown in Table 3.

The first algorithm using only contact information had a sensitivity of 41.9 percent and a specificity of 63.4 percent. These are the measures based upon chance and ignorance and are viewed as the lowest obtainable measures. The sensitivity jumped to 89.4 and 90.9 percent when using the other algorithms, which exploited information in the existing diagnosis-prescription pairs. The not referred to more complicated algorithms only insignificantly improved the prediction (we reached sensitivity rates of 92.6 percent). Consequently, we apply Occams razor and prefer the simple model to the more complex. The figures are found to be clearly acceptable when compared to other applied classification models (e.g. in the binary classification described by Linoff and Berry [11] that reached 85%).

Discussion

Retrospective explanatory linkage of about 90 percent of prescriptions to diagnoses is feasible by statistical methods if there is at least one diagnosis per contact. We were surprised that a simple algorithm performed almost as well as complex algorithms. We had expected that, because patients with single diagnoses are not representative of patients with multiple diagnoses, the same would be true of single-diagnosis contacts versus multiple-diagnoses contacts. On reflection, in light of our results, we realize that complex patients have many single-diagnosis contacts, so prescribing probabilities based only on single-diagnosis contacts can apply to complex patients.

By the same token, we were disappointed that no algorithms achieved sensitivities over 95 percent. Why were some matches so hard to correctly classify as links? The main reason is that some medications were not prescribed during single-diagnosis contacts. For example, aspirin was not prescribed during any single-diagnosis contact. This meant we had nothing on which to base a prescription probability. If the sample size had been larger, more rarely prescribed medications would have been included in the single-diagnosis contacts.

The sensitivities of our method may not be generalizable to linkages involving records of younger patients. Our patient series was deliberately selected to be a more challenging test of the method because patients over 70 often have multiple diagnoses and prescriptions [12].

Also, if participating GPs recorded diagnoses more accurately than average GPs would, then the sensitivities are higher than would be expected with a more representative sample. On the other hand, our “gold standard” cannot be expected to be error proof because such errors produce some random noise that would cause slight underestimates of the sensitivity.

Conclusion

Our general conclusion is that the statistical linkage method is feasible and applicable to other patient groups with less complicated contacts to the GPs.

Although the simple models prove the general behavior among the GPs, the future perspective lies in developing more sophisticated statistical methods as well as experiencing and evaluating the utility of applying software for quality assurance.

A variety of strategies can be taken to improve the linkages:

1) Using combinations of diagnoses that were recorded together during one contact, rather than assessing each diagnosis separately.
2) Using combinations of prescriptions from one contact, rather than assessing each separately.
3) Using combinations of contacts from one patient; a historical or diachronic view.
4) Estimating prescription probabilities separately for different age and sex groups; for example, aspirin among young people is much less likely to be for cardiovascular prophylaxis than among the elderly.
5) Instead of using random numbers to resolve ties, use a library of a priori estimates of probabilities of all rational pairs of ICPC and ATC codes.
6) Finally, other types of more complicated models can be pursued (e.g. artificial neural network); also with the intention of establishing further judgment of the simple models.

If statistical methods can be developed that accurately produce retrospective explanatory linkage, they should be incorporated into computer systems for general practice management. This will save GPs time in data entry and yet enable them to learn how the frequencies of their decisions compare with those of their peers and with guidelines. The result will be that their computers become tools for quality assurance.

References

4. www.medcom.dk.

Correspondence to: Karsten Boye Rasmussen
Organization and Management
University of Southern Denmark
Campusvej 55, 5230 Odense M
Denmark
E-mail: kbr@sam.sdu.dk