electronic Journal of Health Informatics
http://www.ejhi.net
2006; Vol 1 (1): e5

Inferring 'Therapeutic States' of Patients from Community Electronic Prescribing Data

Jim Warren[1], Jan Stanek[1], Svetla Gadzhanova[1], Ivan Iankov[1], Gary Misan[2]
[1] *Advanced Computing Research Centre, University of South Australia*
[2] *Spencer Gulf Rural Health School*

Abstract
We set out to devise a method for analysis of chronic disease therapeutic decision making with specific emphasis on practice patterns across multiple consultations and providers within a single community-based practice. We examine treatment by abstracting each patient's therapeutic state *at any given time as a vector of* n *Boolean state variables, each representing a key decision in the domain under examination. We illustrate the method where the state variables are inferred from electronic prescribing data for treatment of hypertension at a rural practice. We find that graphs of therapeutic state transitions, at various levels of granularity, can provide an overview of prescribing practice or help to identify cohorts of patients that warrant further examination. The graphs, however, are sensitive to heuristic interpretation of the data. A direction for further research is to identify the principles for inference of therapeutic state that are adequately sophisticated for accurate classification of cases and yet interpretable for clinical audit.*

Key Words: Machine Learning, Family Practice, Chronic Disease, Therapy

1 Introduction

There is a positive trend in computer usage by General Practitioners – with the help of financial incentives by government, the proportion of Australian GPs writing prescriptions with the aid of a computer increased from around 50% in 1999 to more than 90% in 2004 (General Practice Computing Group, 2004 cited by Harvey et al., 2005). A survey shows that electronic patient records in general practice could be considered valid, complete and accurate for 99.7% of the prescribed items and for 98.1% of the consultations (Hassey, Gerrett, & Wilson, 2001). Moreover, a systematic literature review of electronic patient record data in primary care (Thiru, Hassey, & Sullivan, 2003) reveals that prescribing data is generally of higher quality than diagnostic and lifestyle data. Hence, electronic prescribing data provides a promising basis for analysis of community medicine, and notably one that is close-at-hand for use at a local level.

Chronic disease management represents a challenge and an opportunity for the appropriate application of IT. It is a challenge because tracking the management of an individual with a chronic condition requires following that patient over long periods of time, across multiple health providers and often between different health sectors (e.g. GP, specialist, public hospitals, community health, allied health, aged care). It is an opportunity because of the prevalence of chronic disease to the community and its high cost. The leading causes of ill health and disability in the Australian population are chronic non-communicable, preventable diseases that relate to the known common risk factors of smoking, nutrition (especially obesity), alcohol consumption, lack of physical activity, high blood pressure and high cholesterol. Over 70% of the burden of illness and injury experienced by the Australian population is associated with cardiovascular disease, cancers, injuries, mental health problems, diabetes mellitus and asthma (Department of Health and Ageing, 2005). Chronic conditions presently comprise the major health burden in developed countries with developing countries trending rapidly toward this, as well. Non-communicable conditions (including mental disorders) accounted for 59% of total mortality in the world and 46% of the global burden of disease in 2000, and chronic conditions are projected to account for 78% of the global disease burden in developing countries by 2020 (World Health Or-

The *electronic Journal of Health Informatics* is an international journal committed to scholarly excellence and dedicated to the advancement of Health Informatics and information technology in healthcare. ISSN: 1446-4381

ganization, 2005). In the US, chronic diseases account for 70% of all deaths, and the medical care costs of people with chronic diseases account for more than 75% of medical care costs (National Center for Chronic Disease Prevention and Health Promotion, 2004). Thus there is an opportunity in that even small improvement in chronic disease management could have significant impact in population wellness as well as on overall health care cost.

We set out to devise a method aimed at analysis of chronic disease therapeutic decision making with specific emphasis on practice patterns across multiple consultations and providers but within a single practice. The goal is to identify treatment patterns that characterize how patients are handled over time with an eye toward practice improvement. Computation on aggregate records can readily reveal excessive referral to specific testing facilities or trends toward excessive prescription of antibiotics; for management of chronic diseases such as hypertension, however, there are numerous classes of perfectly appropriate drugs that may be prescribed individually or in combinations. It is indirect and insufficient to assess prescribing of, say, beta-blockers, in terms of overall frequency. Guideline-adherent prescribing is only assessable in terms of the order in which therapies are attempted, and with an ordering that varies depending on complications (i.e., other patient conditions outside the condition that is the subject of the treatment guideline, known as *comorbidities*).

Our method is based on events defined in terms of which drugs have been prescribed as shown in the electronic health record (EHR) of the practice's prescribing system. In fact, each prescribing event in the EHR defines two events – the start of the prescription, and a second event when the prescribed supply should have run out if the patient is adherent to the dosing directions and gets all 'repeats' (refills) of the prescribed medication. These prescribing-based events stand in contrast to quantitative measurements (such as physical measurements of the patient) that could be characterized as parameter values (e.g., high or low), gradient of change of parameters (e.g., increasing or decreasing) and rate of change of parameters (e.g., fast or slow) as in KBTA (Knowledge-Based Temporal Abstraction) (Shahar, 1997). Another feature of the proposed method is stipulated by the type of general practice medicine data – on the contrary to most of the existing methods of temporal abstraction in medicine that work with data monitored within short periods of time and taken on regular bases (Larizza, Ballazzi, & Riva, 1997; Miksch, Horn, Popow, & Paky, 1996), we need to deal with long-term, irregularly-spaced data – a patient could have several visits within a short time followed by a long period of time (several months) without any action recorded. While these differences are not absolute (e.g., one can look at the dosage level of a particular medication to define KBTA-type parameters), we believe the difference is fundamental and use it to define a different approach to analysis of EHR data.

In the following sections we describe and illustrate application of our therapeutic state transition model. Application of the model requires heuristic interpretation of the prescribing events present in the EHR. We illustrate sensitivities of state-transition output to heuristic decisions. We then provide some brief examples of the use of the modelling results to focus clinical practice quality improvement efforts. We conclude with a discussion of our short and long-term research goals.

2 Method

A major goal of our research is to provide General Practitioners (as community based prescribers) with an introspective view on how their own practice is managing the treatment of selected groups of patients. Our basic concept is to abstract information from EHRs and transform it into series of states and transitions in order to identify how patients' therapy matches against the states and transitions described by a chosen guideline, which will allow the clinicians to draw their own conclusions about appropriateness of actual practice. A criterion for choosing a guideline is to identify an evidence-based one, containing recommendations for optimal treatment of the group of patients of interest. It could be then modified to suit the local context. Our model will enable GPs to see how patients flow through pathways implied by the guideline and whether some treatment patterns fall outside of its recommendations.

Based on the chosen guideline, a number of Boolean (yes/no) variables are defined to represent major treatment options. For our purposes we denote that a therapeutic state variable will be synonymous with a specific drug group (such as Diuretics or Beta-blocking agents), where drug groups are identified from the guideline. From EHR prescribing data we compute the state of a patient on each given day in an investigated period as the combination of state variables (i.e., state representing which of the drug groups is in use for that patient on that day). Thus, the patient history will be a sequence of states in which the patient has been during the investigated period and state transition will exist where the state value changes. A state-transition diagram will be the visual representation of the identified histories of the group of patients of interest, offering a straightforward way for the practice's management of a patient cohort over the investigation period to be reviewed by the medical practitioners. Certain state-transition paths that appear to differ from the "usual" practice (i.e., that do not appear to be the main stream of the practice and are not easily reconciled with the clinical practice guideline) are brought to the attention of the GPs for further consideration.

State-Transition Model

The *coverage period* of a prescription is defined by the estimated duration of the prescription (time in days that the prescription should last the patient in light of dose, instructions, pack size and number of repeats) expressed as a date range.

To approach the formal definition of therapeutic state let us say: n is the

number of observed drug groups (i.e., number of therapeutic state variables); *m* is the number of observed patients; and *p* is the number days in the investigation period.

Let t_k denote the k^{th} day of the analysis period, $1 \leq k \leq p$.

Define $g_{ij}^{t_k}$ as:

$g_{ij}^{t_k}$ =0, if the j^{th} patient is not covered by a prescription from the i^{th} drug group at t_k

$g_{ij}^{t_k}$ =1, otherwise

where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Let the value of the therapeutic state of patient *j* at t_k, denoted $S_j^{t_k}$, be the Boolean vector

$$S_j^{t_k} = [g_{1j}^{t_k}\, g_{2j}^{t_k} \ldots g_{nj}^{t_k}]$$

One can see that there are 2^n mathematically possible values of a patient's therapeutic state. It is useful to refer to patient *j* being in the "zero state" at time t_k if $g_{ij}^{t_k} = 0$, $\forall i = 1,2,\ldots,n$.

Where the value of $S_j^{t_k} \neq S_j^{t_{k+1}}$, we say patient *j* has made a *transition* at time *k*+1. The sequence of state values held by a patient defines that patient's *path*.

Heuristic Interpretation of Electronic Prescribing Data

While the EHR data is likely to provide a highly accurate record of GP prescribing decisions for the prescribers at the practice under consideration, there are a number of sources of uncertainty external to the database for interpretation of the data as real-world events:

Non-compliance

It is not certain that the patient has filled the prescription, or, if so, takes the medication as directed and for the full duration specified;

Other prescribers and drug supply sources

The patient may see another GP, a specialist or get a prescription from a hospital. These possibilities are somewhat diminished in an isolated rural practice. Moreover, the patient may receive drug supply as sample packets (sometimes GPs give drug company samples to patients rather than have them go to waste) or by using someone else's medication (e.g., their spouse's);

Implicit stop/change events

At any time, the GP or patient may decide to stop taking a medication, which will not have a record in the EHR. Moreover, the EHR does not generally indicate if a new medication prescription is in addition to an ongoing one or represents a replacement; and

Scheduling vagaries

A patient may arrive some amount of time early or late to get a new prescription for an ongoing therapy with no specific relationship to a therapeutic decision.

As such, the prescribing data does not provide as crisp an indication of therapeutic decisions and the patient's actual therapy as one might expect. Several heuristics are helpful to make the raw state-transition data more interpretable. Each can be parameterised, and in the next section we will see that analysis results vary considerably on these parameter values.

Overlap

With respect to ongoing use of one class of medication, a patient can reasonably accrue a small supply of prescription 'backlog' such that if they are late for a subsequent visit, it is appropriate to consider the earlier excess supply. However, there is a limit to the usable backlog, notably because prescriptions eventually expire. As a special, extreme case of overlap we ignore all but one prescription in the case of multiple prescriptions for the same drug on the same day as spot validation has shown these to often be due to errors in printing rather than accurate data.

Short

A very brief time at a state probably does *not* indicate a specific therapeutic regime, but instead is simply the cusp of a change from one longer-term therapeutic state to another. Below a threshold duration, a Short state is absorbed into the previous state.

Zero

A brief (or even not-so-brief) time with no medication, if bracketed by two periods of active therapy is probably a case of inadvertent late attendance to a represcription visit or a result of undocumented excess drug supply (or non-compliance) at levels that are not worthy of further scrutiny. The duration of the zero state is absorbed into the prior state.

Temp

A brief dip into a state that is bracketed by another particular state before and after may be reasonable to ignore, even for periods somewhat longer than a Short state.

In addition to the heuristics, there is a question of how many patient transitions constitute the minimum to highlight for analysis (in particular, in a graphical state transition display). One school of thought is to include every single transition, but a therapeutic path that is not repeated is unlikely to represent GP intent, and as such has minimal value to highlight for quality improvement. We find that displaying only transitions that have been repeated on the order of 5 to 10 times (for our size of data set) is most useful. Obviously, the choice of specific values for the above heuristic parameters constitutes a significant decision, each involving assumptions about what is, and is not, important. The effects of these heuristic parameters are illustrated in the next section.

3 Illustration of the Method

We illustrate application of our model by investigating data from a rural medical practice in South Australia with two full-time GPs, where six years of EHR data are available providing approximately 70,000 prescriptions on 4,000 patients. We are particularly interested in the group of patients diagnosed with diabetes mellitus and hypertension at some point in the EHR data, because diabetes affects 7.4% of Australians over the age of 25 (Dunstan et al., 2002) and hypertension is the most common co-morbidity (affecting 20-60% of the diabetic patients), causing 86% of deaths in patients with diabetes (Wingard & Barrett-Connor, 1995). In fact, the practice was quite aware of the importance of tightly monitoring diabetes patients and had assigned a nurse to attend specifically to the tracking of this group. Under these conditions it is unsurprising that we found the diabetes cohort to be receiving very well-considered therapy and timely follow-up (a diabetes patient is sent a follow-up letter if they miss a six-monthly appointment). For purposes of the present paper it is more interesting to construct a state-transition model for patients who were coded in the practice EHRs as having hypertension but *not* diabetes mellitus (529 patients). This group, not receiving systematic reminders, provides a richer illustration of situations that may arise without intensive human attention to the management of each case. Moreover, it should be considered that it is not uncommon for a patient to develop and/or be detected with diabetes during the course of their long-term hypertension management.

For the purposes of our analysis we identified a synthesis of Australian and international guidelines (American Diabetes Association, 2004; Chobanian, 2003; Gilbert, Jasik, DeLuise, O'Callaghan, & Cooper, 1995; Smith, 2003; Department of Health, 2001; Plouin, 1997) for the management of patients with combined diabetes and hypertension. We abstracted six relevant groups (therapeutic state variables) of therapeutic agents for hypertension in diabetes, defined in terms of Anatomical Therapeutic Chemical (ATC) classification (WHO Collaborating Centre for Drug Statistics and Methodology, 2003). (Noting that these therapeutic agents are also the ones applicable for hypertension patients *without* diabetes, but with less restrictions in terms of which are most strongly recommended for initial therapy.)

A: ACEi (Angiotensine Converting Enzyme inhibitors) and ARBs (Angiotensine Receptor Blockers) (ATC: C02EA)

B1: Beta-blockers (ATC: C07, especially discerning selective beta-blockers: C07AB)

B2: Diuretics (ATC: C03AA – thiazides and C03C – loop diuretics)

B3: Non-dihydropyridine calcium channel blockers (ATC:C08CX, C08DA01, C08DB01)

C: Dihydropyridine Calcium Channel Blockers (DCCBs; ATC: C08CA)

D: Alpha blockers, hydralazines and clonidine (ATC: C02DB, C02CA, C02AC)

We also include in the analysis combination products that represent more than one of the groups – ATC: C07BA, C07BB, C03AB, C08GA01, C09BA, C09BB and C09DA.

Applying Heuristics to Achieve Therapeutic State-Transition Results

Table 1 describes the number of nodes (distinct states) and arcs (distinct state transitions) found at various heuristic parameter values (see previous section on *Heuristic Interpretation of Electronic Prescribing Data*) for a two-year analysis period for patients diagnosed with hypertension and not diagnosed with diabetes. Figures 1-3 show state transition graphs for various specific rows of table 1. In the figures, each node shows a therapeutic state (combination therapy is represented with symbols separated by commas) and below the state we have included number of patients in this state at the beginning of the investigated time period followed by the number of patients at the end of the period. Arcs are annotated with the number of transitions between two states and (in parentheses) total days spent in the destination state as well as the number of distinct patients making these transitions. Init_In ("Initially in") is defined as a state where the patient has had no prescription for a drug group of interest (the six groups of

Heuristic Parameters (days)				**Transition granularity** (min. transitions to display)							
				≥1	≥5	≥10	≥15	≥1	≥5	≥10	≥15
Overlap	*Short*	*Zero*	*Temp*	# nodes in graph				# arcs in graph			
0	0	0	0	46	26	15	14	246	90	49	41
30	10	30	30	45	21	11	11	225	55	30	21
60	20	60	60	41	15	11	10	197	37	23	19
90	30	90	90	40	17	11	9	190	32	19	11

Table 1. The effects of heuristics and graph display granularity on number of nodes and arcs in the therapeutic state transition graph – a selection of illustrative values for hypertension prescribing to non-diabetes patients in one general practice.

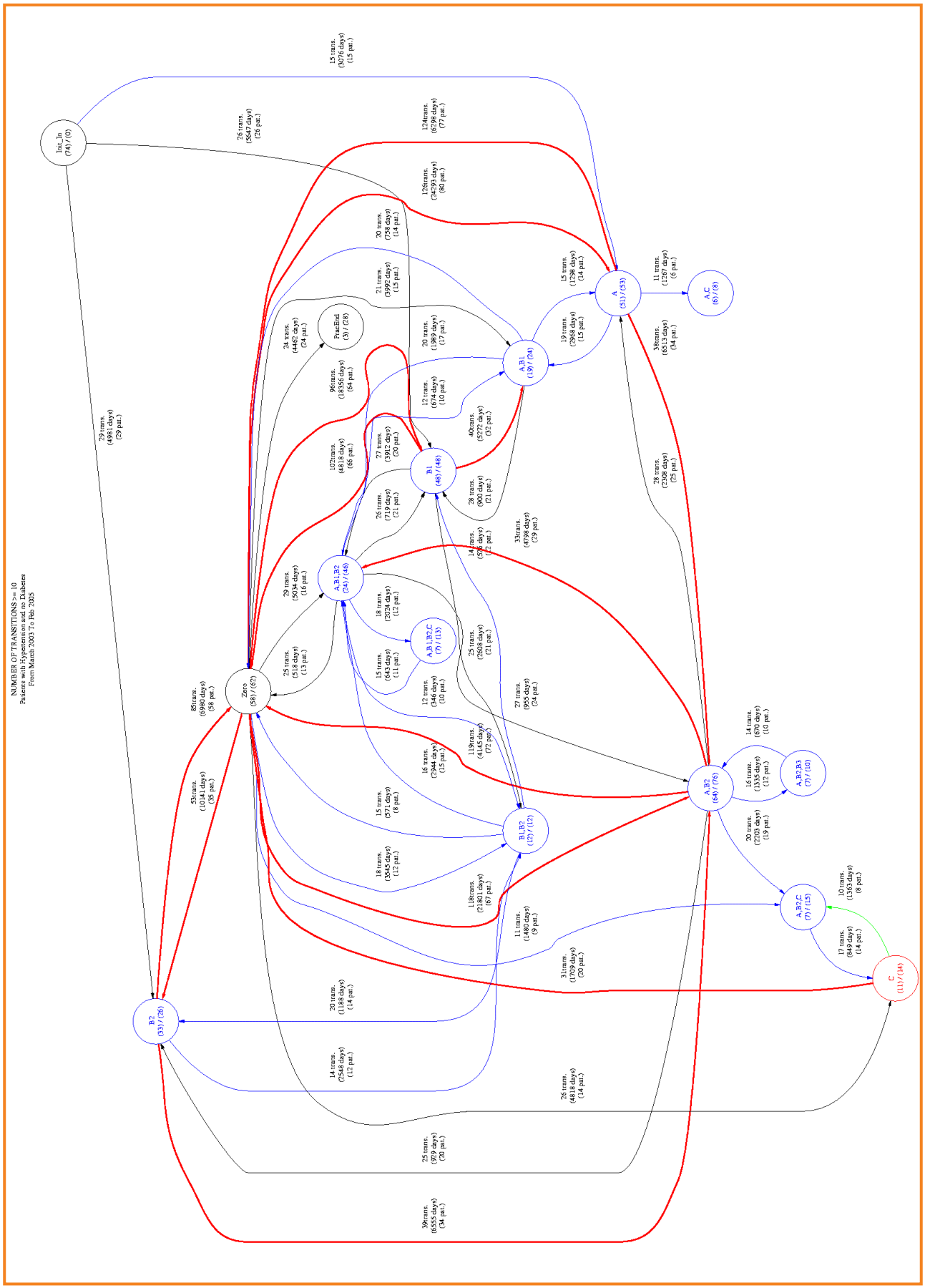

Figure 1. State-transition graph for patients with hypertension and no diabetes for heuristic values 0, 0, 0, 0 (see table 1), displaying arcs with ≥10 transitions.

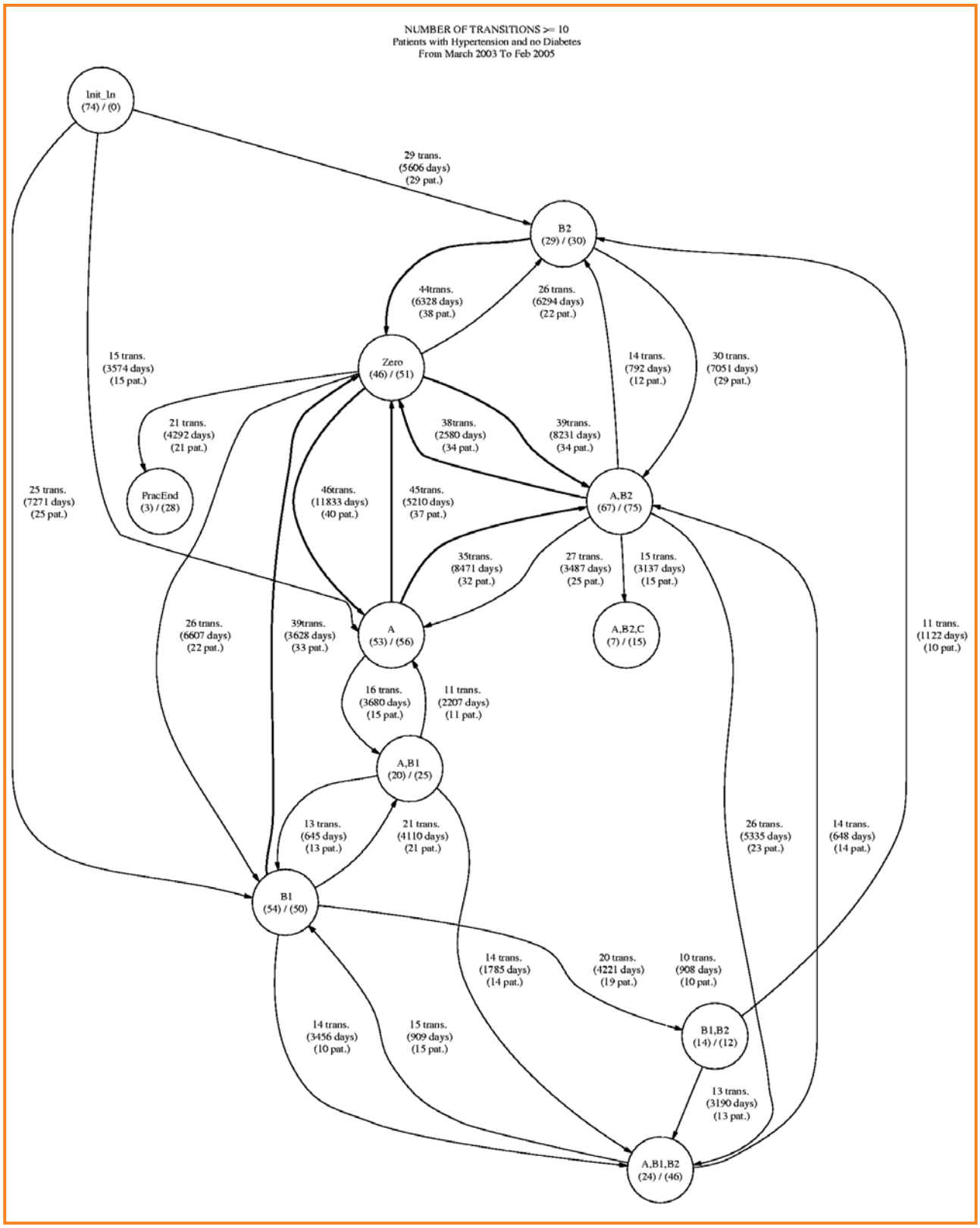

Figure 2. State-transition graph for patients with hypertension and no diabetes for heuristic values 30, 10, 30, 30 (see table 1), displaying arcs with ≥ 10 transitions; an intermediate level of granularity.

antihypertensives) in the EHR history, while still having encounters with the practice. We consider at least 100 days of contact with the practice before antihypertensive therapy to constitute Init_In. Zero state implies no active therapy for a period of *n* days (where *n* is as per the Zero heuristic parameter value).

Using State-Transitions in Practice Quality Assurance

It is notable that many transitions and some states drop away as we right

monitoring tests are conducted. This does not mean that these cases are not the most worthy of continued audit attention. Conversely, we have found that some specific fairly simple cases are quite useful indicators for clinical audit: for instance, the C state (C only, with no combination) in patients with diabetes *and* hypertension is relatively difficult to reach when following practice guidelines – some of these patients were actually diagnosed with diabetes while receiving the C-state therapy, and others were found to be relatively worthy of review. We are currently in the process of further tuning our set of alert states and transitions and compiling the necessary screening test results or other clinical conditions which ameliorate these alerts.

4 Discussion and Future Research

We have described and illustrated a method of analysis based on therapeutic state which identifies patients of interest for clinical practice audit without resorting, in the first instance, to features of the patients per se, but look-

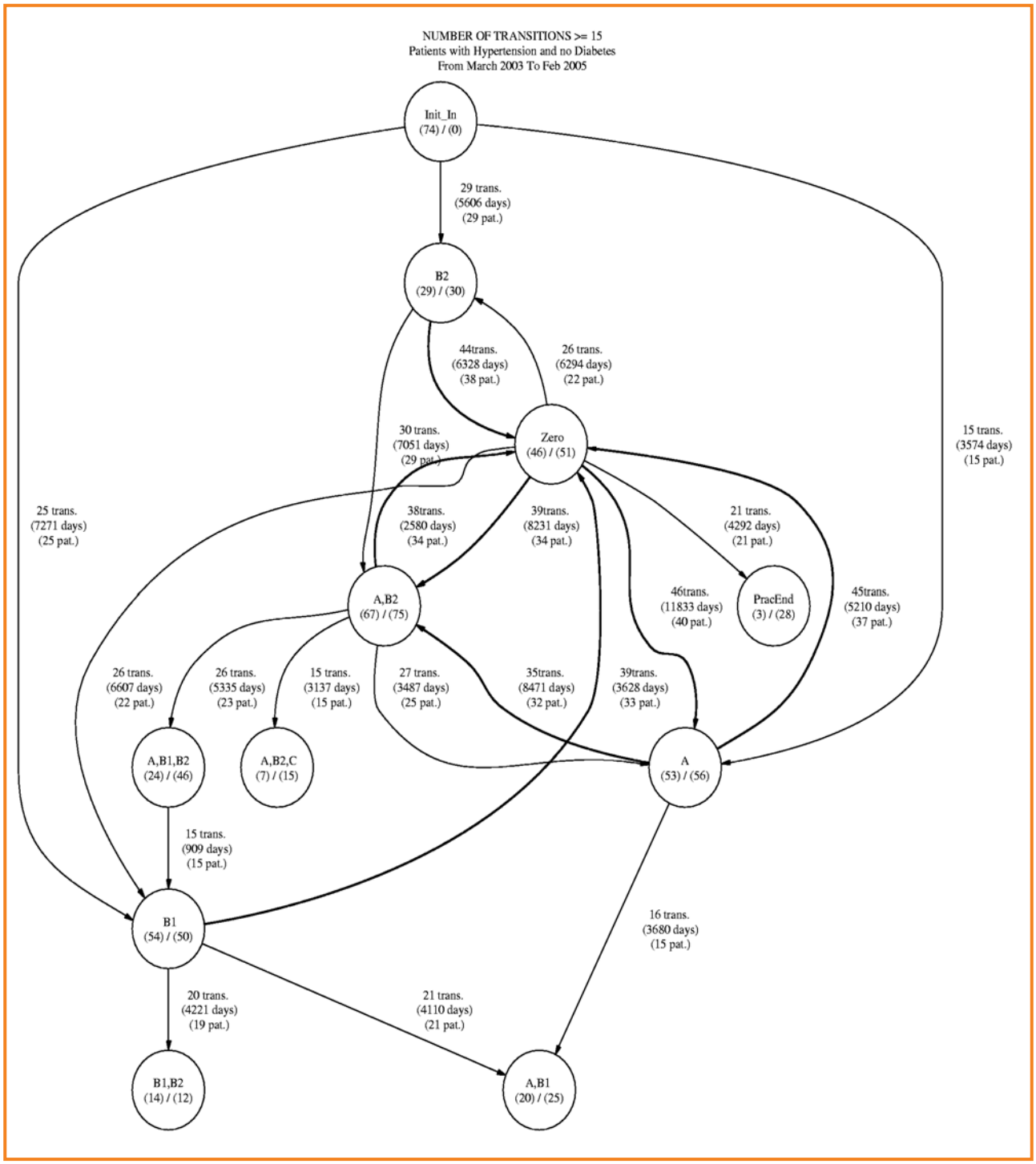

Figure 3. State-transition graph for patients with hypertension and no diabetes for heuristic values 30, 10, 30, 30 (see table 1), displaying arcs with ≥15 transitions; a good choice for viewing the overall practice pattern.

ing exclusively at the series of management actions taken on those patients over time. Others have explored the concept of applying machine learning/data mining techniques to compare clinical guidelines to actual clinical decision making (e.g. Sboner & Aliferis, 2005), but we believe our technique has a unique level of practicality for examination of chronic disease management in community-based practices. In addition to providing a tool for post hoc quality assurance and ongoing process improvement, the same method can be used as a test-bed for future implementation of decision support alerts or reminders. While therapeutic state alone identifies classes of patients that are surprisingly rich in interesting cases, we believe we can reduce the false-positive rate by adding more conventional domain-knowledge rules about required tests and other conditions.

Our immediate further research plan involves assessing a state-based method, augmented with a limited number of conventional screening tests, in terms of sensitivity and specificity in detecting cases of relatively interesting cases. We will measure interestingness in terms of GP assessment of relative merit for review and also in terms of favourable reaction to the prospect of a decision support alert or reminder at the time of a state-transition event.

A longer-term research question is how to move from our (clearly important) set of heuristics to an extended model for inference of therapeutic state. We would like to use a measure of uncertainty explicitly in the representation of patient state, rather than sharp cut-off values based on threshold numbers of days. A probabilistic model has some merit, e.g. to say that patient in the Zero state for 60 days after A,B1 state has a 50% chance of being out-of-supply of their prescribed drug. A good case can be made, however, for a Fuzzy Set Theory representation of state, e.g. a patient in the Zero state for 60 days after A,B1 state has a 0.5 degree of membership in the set of patients that are interesting to review. However, any complex model of therapeutic state runs the risk of being unusable as an audit tool because of the increased difficulty of explaining, from a medico-legal context, why a patient was or was not highlighted for review.

References

American Diabetes Association. (2004). Hypertension management in adults with diabetes. *Diabetes Care*, 27:S65-S67.

Chobanian, A. (Ed.). (2003). *The Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure.* US Dept. Health and Human Services. National Institutes of Health Publ. No. 03-5233.

Department of Health. (2001, December 14). National service framework for diabetes: standards. Retrieved December 2005, from http://www.dh.gov.uk/PolicyAndGuidance/HealthAndSocialCareTopics/Diabetes/fs/en

Department of Health and Ageing. (2005, February 24). National strategy for an ageing Australia. November 2005, from http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/ageing-ofoa-agepolicy-nsaa-nsaabk7.htm

Dunstan, D., Zimmet, P.,Welborn, T., Cameron, A., Shaw, J., de Courten, M., Jolley, D., & McCarty, D. (2002). The Australian diabetes, obesity and lifestyle study (AusDiab) – methods and response rates. *Diabetes Research and Clinical Practice*, 57, 119-129.

General Practice Computing Group. (2004, May 27). Practice incentives program. Retrieved May 2005, from http://www.gpcg.org/topics/pip.html

Gilbert, R., Jasik, M., DeLuise, M., O'Callaghan, C., & Cooper, M. (1995). Diabetes and hypertension – Australian diabetes society position statement. *Medical Journal of Australia, 163*, 372-375.

Harvey, K., Vitry, A., Roughead, E., Aroni, R., Ballenden, N., & Foggotter, R. (2005). Pharmaceutical advertisements in prescribing software: an analysis. *Medical Journal of Australia*, 183(2), 75-79.

Hassey, A., Gerrett, D., & Wilson, A. (2001). A survey of validity and utility of electronic patient records in a general practice. *British Medical Journal*, 322, 1401-1405.

Larizza, C., Bellazzi, R., & Riva A. (1997). Temporal abstraction for diabetic patients management. *Artificial Intelligence in Medicine. Proc AIME-97*, 319-330, Springer.

Miksch, S., Horn, W., Popow, C, & Paky F. (1996). Utilizing temporal data abstraction for data validation and therapy planning for artificially ventilated newborn infants. *Artificial Intelligence in Medicine* 8(6), 543-576.

National Center for Chronic Disease Prevention and Health Promotion. (2004, October 15). Chronic disease overview. Retrieved November 2005 from http://www.cdc.gov/nccdphp/overview.htm

Plouin, P-F. (Ed.). (1997). *Diagnostic et traitement de l'hypertension arterielle essentielle de l'adulte de 20 a 80 ans.* ANAES (Agence Nationalle d'Accreditation et d'Evaluation en Sante).

Sboner, A., & Aliferis, C. (2005). Modeling clinical judgment and implicit guideline compliance in the diagnosis of melanomas using machine learning. *Proceedings of the AMIA Annual Symposium*, pp. 664-668.

Shahar, V. (1997). A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90, 79-133.

Smith, A. (Ed.). (2003). *Therapeutic guidelines – cardiovascular.* (4th ed.). Melbourne, Therapeutic Guidelines Inc.

Thiru, K., Hassey, A., & Sullivan, F. (2003). Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal*, 326, 1070-80.

World Health Organization. (2005, July 1). Chronic conditions: the global burden. Retrieved November 2005, from http://www.who.int/chronic_conditions/burden/en/

WHO Collaborating Centre for Drug Statistics Methodology. (2003). Anatomical Therapeutic Chemical (ATC) classification. Retrieved February 2003 from http://www.whocc.no/atcddd/

Wingard, D. & Barrett-Connor, E. (1995). Heart disease and diabetes. In, *Diabetes in America*, Washington, DC, U.S. Govt. Printing Office, 429-448 (NIH publ. No. 95-1468).

Correspondence

Prof. Jim Warren
Department of Computer Science - Tamaki
University of Auckland
Private Bag 92019 Auckland
New Zealand

jim@cs.auckland.ac.nz