Informatics in Primary Care 2011;19:57–63

Leading article

Data-modelling and visualisation in chronic kidney disease (CKD): a step towards personalised medicine

Norman Poh PhD
Research Fellow, Centre for Vision, Speech and Signal Processing (CVSSP)

Simon de Lusignan MBBS MSc MD(Res) FBCS FRCGP
Professor of Primary Care and Clinical Informatics, Department of Health Care Policy and Management, University of Surrey, Guildford, Surrey, UK

ABSTRACT

Background Personalised medicine involves customising management to meet patients' needs. In chronic kidney disease (CKD) at the population level there is steady decline in renal function with increasing age; and progressive CKD has been defined as marked variation from this rate of decline.

Objective To create visualisations of individual patient's renal function and display smoothed trend lines and confidence intervals for their renal function and other important co-variants.

Method Applying advanced pattern recognition techniques developed in biometrics to routinely collected primary care data collected as part of the Quality Improvement in Chronic Kidney Disease (QICKD) trial. We plotted trend lines, using regression, and confidence intervals for individual patients. We also created a visualisation which allowed renal function to be compared with six other co-variants: glycated haemoglobin (HbA1c), body mass index (BMI), BP, and therapy. The outputs were reviewed by an expert panel.

Results We successfully extracted and displayed data. We demonstrated that estimated glomerular filtration (eGFR) is a noisy variable, and showed that a large number of people would exceed the 'progressive CKD' criteria. We created a data display that could be readily automated. This display was well received by our expert panel but requires extensive development before testing in a clinical setting.

Conclusions It is feasible to utilise data visualisation methods developed in biometrics to look at CKD data. The criteria for defining 'progressive CKD' need revisiting, as many patients exceed them. Further development work and testing is needed to explore whether this type of data modelling and visualisation might improve patient care.

Keywords: automatic data processing, chronic disease, diabetes mellitus, longitudinal studies, personalised medicine, renal insufficiency

Introduction

Decision making in the context of the 10-minute consultation can be challenging, and the advent of personalised medicine might make the task even more complex. The concept of personalised medicine had been borne out of a combination of: improvements in our knowledge about the human genome; understanding of biomarkers and other developments in basic science; and advances in computational power.[1] Personalised medicine is generally described as the dividing of patients into groups on the basis of their biological or genomic make-up; and personalising management regimes for these groups. However, as clinicians we sit with individual patients and need to develop individual strategies for every patient, and have some insight into whether they are getting better or getting worse with often very few biochemical measures at our disposal; and, although our computerised medical record systems may hold a decade

or more of readily available data;[2,3] though it is not necessarily easy to interpret. Visualising longitudinal patient data may help their clinician observe the interplay between markers of disease and therapy on an individual patient basis. Longitudinal data held in electronic patient records may provide insight into whether people are responsive or not to therapy.

Chronic kidney disease (CKD) is an important condition which affects between 5% and 10% of the population;[4] and interventions which can affect outcome, particularly strict control of systolic BP,[5] can be implemented in primary care. It is important because it is associated with an increased risk of major cardiovascular events and death; broadly similar to the excess mortality and morbidity associated with diabetes.[6] There is clear guidance how to manage CKD in the adult population,[7,8] however, GPs find the guidance challenging to implement,[9] and lack confidence in managing CKD compared with other comorbidities.[10] Difficulties include that renal function, primarily measured using estimated glomerular filtration rate (eGFR), fluctuates markedly; this because it is derived from serum creatinine, primarily a product of muscle breakdown and protein containing food. The national guidance for management of CKD in adults suggests that a decline in eGFR of >5 ml/min in one year or a decline of >10 ml/min in five years should be regarded as rapid decline and referred for specialist care. Data from the Quality Improvement in Chronic Kidney Disease (QICKD) trial[11] suggests that the decline seen year on year in the population in eGFR is between 0.4 and 0.7 ml/min per year. Figure 1 shows the decline in eGFR during adult life and superimposed are the rates of decline where referral is recommended.

However, looking at the population as a whole there is considerable fluctuation in creatinine, with much of this fluctuation exceeding the decline in eGFR of >5 ml/min in one year or >10 ml/min in five years recommended as thresholds for referral.[12] For example, in the QICKD trial we found that 31% (16 587/53 322) people with two eGFR readings more than a year apart had their creatinine fluctuate by over 5 ml/min in one year (Figure 2); clearly it would not have been feasible to refer them all.

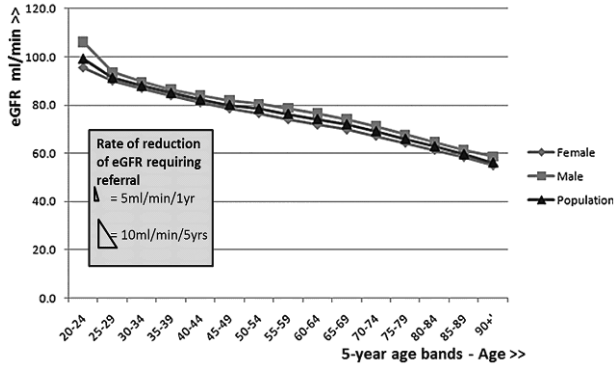

Figure 1 The rate of decline in eGFR with increasing age for males, females and the population; superimposed are triangles to represent accelerated decline where referral is recommended. From QICKD study baseline data (*n*= 50 331, people with CKD)

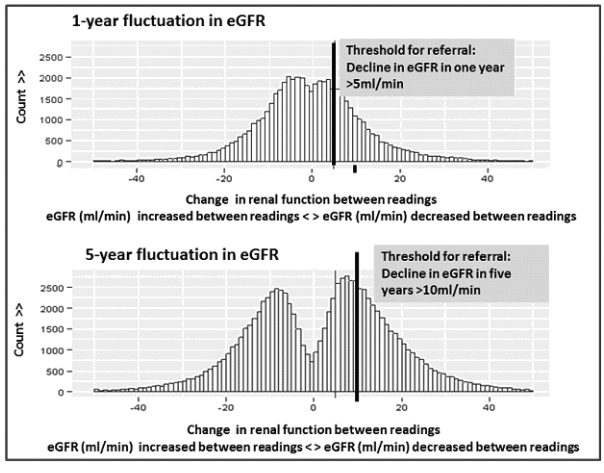

Figure 2 Fluctuation between minimum and maximum eGFR over one and five years; the vertical bold lines indicate a decline of over 5 and 10 ml/min in eGFR, which should be referred to specialist care

Unfortunately combining our knowledge about the population trend and how eGFR fluctuates in the population does not help greatly with the management of individual patients. In the clinical scenario the responsible clinician has to decide whether the change in eGFR is part of normal fluctuation or not for the individual patient in front of them. We carried out this pilot study to show whether we could use visualisation techniques to display trends in renal function in individual patients; and indicate whether a new value is part of normal fluctuation or an abnormal finding.

Method

The aim of the pilot is to demonstrate methods of data visualisation to display; (1) fluctuation among individual patients within a population with apparently smooth change; and (2) create pilot outputs from routine data that can be used to help manage individual patients; using methods which can be readily made machine processable. Longitudinal displays of routine data have been proposed as a method that might help to improve care.[13] We worked with a small panel of clinical experts in CKD, nephrologists, researchers and general practitioners, to understand how improved display of data might improve clinical management.

The expert panel confirmed the need for a data-presentation methodology that is personalised and patient-centred. We have set out in the introduction how current methodologies, based on analysis at the population level, are not adequate because the trend in each patient is aggregated; the methodological challenge is to produce an improved version.

We transferred techniques developed from biometrics to medical science because both domains share the following similarities: they deal with signals generated from biology processes; both types of data can be 1D, 2D, 3D or 3D+time; both type of data are longitudinal in nature, often extending over many years; the types of data samples are collected in open environments, or *in-vivo*, that is, the experimenter has minimal control over the experimental setting. The last characteristic makes the data modelling problem extremely challenging.

The biometric methods applied to the data modelling problem in CKD are as follows:

- estimate the biometric performance over time – in the case of CKD we applied this approach to eGFR as the key measure of renal function[14]
- the model parameters: primarily how eGFR fluctuates with age, is calibrated using a regression model that hypothesises what is acceptable change over time[15]
- the method produces an output at the individual patient level which might ultimately be capable of supporting improved medical decision making[16]
- the method that handles missing values in data modelling[17]
- a method that models a hypothesis (person identity) taking contextual covariates into account.[18]

The assumption in the method is that that the biomedical signal is generated by a steady process that is corrupted by daily biologically induced fluctuation and other activities unaccounted for (collectively considered as 'noise'). Creatinine production is known to be generated by muscle breakdown and its serum level is related to muscle mass; the steady process. Generally as renal function declines blood levels of creatinine rise; again a steady process of creatinine rising and eGFR falling with increasing age (Figure 1). However, there are a number of factors which can cause creatinine to change; and create 'noise', or fluctuation. Eating protein based meals and increased cell breakdown (due to illness or treatments for cancer) can increase serum creatinine (SCr); and eating little and reducing muscle mass (e.g. an amputation) can reduce it. eGFR is related to the inverse of SCr, so as SCr rises with deteriorating renal function so eGFR falls.

The data for the QICKD trial are held on a secure MySQL[19] database. Customised data extraction queries were developed to provide test variables for this exercise. These queries mimic those that might be used to extract data from live clinical systems rather than from a research database.

We fitted a regression line to the eGFR values available on a per patient *basis*, and for the entire population. This regression line has the effect of smoothing the 'noise' or fluctuation in eGFR. The regression line also enables 95% confidence intervals to be plotted, giving some indication for that individual patient whether their fluctuation is within an acceptable limit based on previous change or potentially greater. We plotted fitted regression curves, superimposed with the original data, for a random selection of 20 patients with over 15 years' data.

We then developed a data display data for an individual patient with CKD, and diabetes. We displayed:

- eGFR as a measure of renal function
- glycated haemoglobin (HbA1c) – a measure of the quality of diabetes control (using International Federation of Clinical Chemists (IFCC) units – the nearer to 55 mmol/mol the better the control)
- body mass index (BMI) – a measure of obesity (weight in kg/height in metres squared) BMI over 30 kg/m^2 defines obesity
- blood pressure (BP) – measures as systolic and diastolic (a BP of under 130/80 is recommended)
- prescribed Metformin – this is a first line anti-diabetic drug – which is useful because it helps

reduce appetite but has been associated with impaired renal function [20]
- prescribed other oral anti-diabetic drugs (OAD) – these drugs are effective in controlling diabetes, but by lowering blood glucose and increasing appetite and result in weight gain
- we also plotted the consumption of aspirin, as an example of a cardiovascular medication.

We plotted a regression line and 95% confidence intervals for the variation in eGFR, HbA1c and BMI. We initially attempted this with BP but it involved the display of too many lines. We plotted drug consumption on consumption of tablets or item per 28 days. We weighted up the pros and cons of using a medicines possession ratio, as we have previously used for osteoporosis medication;[21] however, as we may not always know the intended dose of medication we considered that the number of tablets or items per 28 days would be of most value.

Ethical considerations

The QICKD Trial is ethically approved and this study's details are recorded at the International Standard Randomised Controlled Trial Number Register (ISRCTN56023731).[22]

Results

We display how beneath the smooth and linear decline in eGFR (Figure 1) there is considerable noise/fluctuation among individuals. The fitted regression curves which are superimposed with the original data for 20 patients are shown in Figure 3. For example: on the extreme right of the figure is a person for whom we have data from their late 80s through to nearly 100 years of age. During this period the patient's 'smoothed' eGFR rises from 42 ml/min to 56 ml/min. However, their fluctuating eGFR changes from 48 ml/min down to its lowest level 39 ml/min eventually finishing at 58 ml/min. For this individual, although their trend overall was one of improvement, they might have been referred on the basis of their rapidly declining eGFR.

We presented these figures to a panel of CKD experts; they thought this approach offered more than just comparing individuals with the population trend (Figure 1). They felt that the smoothed data made it much easier to observe trends in patients and reported that data presented in this way had the potential to enable an estimate to be made of the annual rate of change of eGFR in an individual. They speculated that if that could be based on their smoothed regression line, and whether this more accurately represented their underlying change in renal function. They felt that individual case examination 'must' be superior to the current approach which is to compare an individual's change with arbitrarily set limits which define an accel-

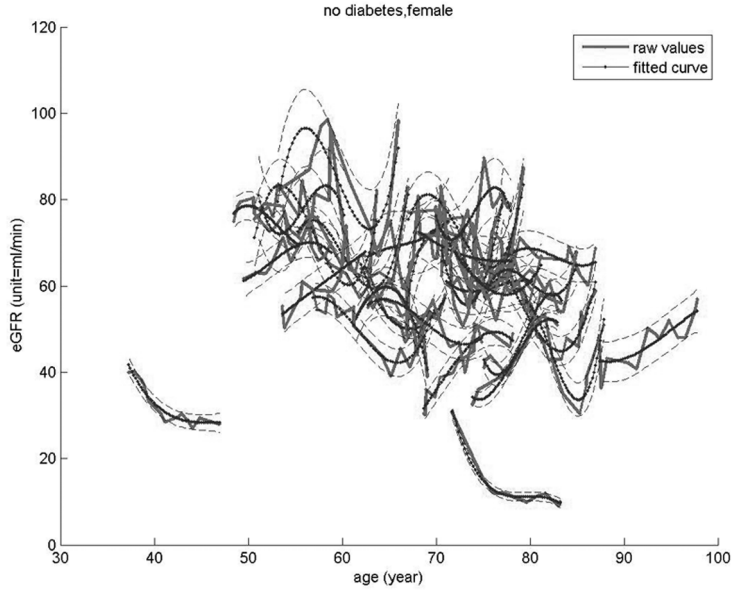

Figure 3 Fluctuation in individual's eGFR. This figure shows raw values (bold fluctuating line) then a regression line (the fitted curve) for each individual, we also display as dotted lines 95% confidence intervals for the regression line. All the patients have at least 10 years data.

erated decline (i.e. >5 ml/min on one year, >10 ml/min in 5 years).

We developed a viewer for individual patient data. This data visualisation displays anonymised, real patient data for one patient drawn from the QICKD trial. This is shown as an exemplar of how machine processable data might be displayed (Figure 4). For this single patient we display data from between the ages of 63 and 81 years. There is only four years' eGFR data, but this shows an initial decline then a levelling off in that decline. Glycated haemoglobin (HbA1c) has risen steadily between age 70 and 78 years, but has plateaued and moved back towards target values over the last 3–4 years. The improvement in HbA1c appears to be mirrored by a change in BMI, a fall in BMI appears to be associated with a fall in HbA1c. Systolic BP generally appears to have been higher in the last decade. The step up in BP is associated with the time when metformin was first prescribed, which in turn may have been associated with a diagnosis of diabetes. One short period on an OAD was associated with a fall in HbA1c, however, the association between falling HbA1c and BMI appears stronger.

Initial opinion from our panel was generally positive. They felt this type of computer generated display might significantly improve care in complex cases. They thought that the change in BMI and HbA1c may have been hard to discern in usual care records, some of which allow listing or graphing of variables but not as many variables, nor as succinctly. No existing providers they were aware of provide regression lines to illustrate the trend, or confidence intervals to give some idea if a value is an outlier. However, they also thought that considerable testing was needed of the combinations of variables needed for effective management. For example, should proteinuria measures always be displayed with eGFR; haemoglobin with aspirin, other anticoagulants, and non-steroidal anti-inflammatory drugs; the need to display blood pressure medications, maybe angiotensin modulating drugs separately; and others.

Discussion

Principal findings

We have successfully demonstrated that routine data in UK practice had significant amounts of longitudinal data. We have also shown that smooth trends in eGFR at the population level are not matched by changes at the individual patient level. Additionally, many patients have changes in eGFR which exceed the national guidance definition of 'rapidly declining' eGFR. Finally, we have created a computer generated visual display of a range of variables and fitted curves (Figure 4); which was well received by a small panel.

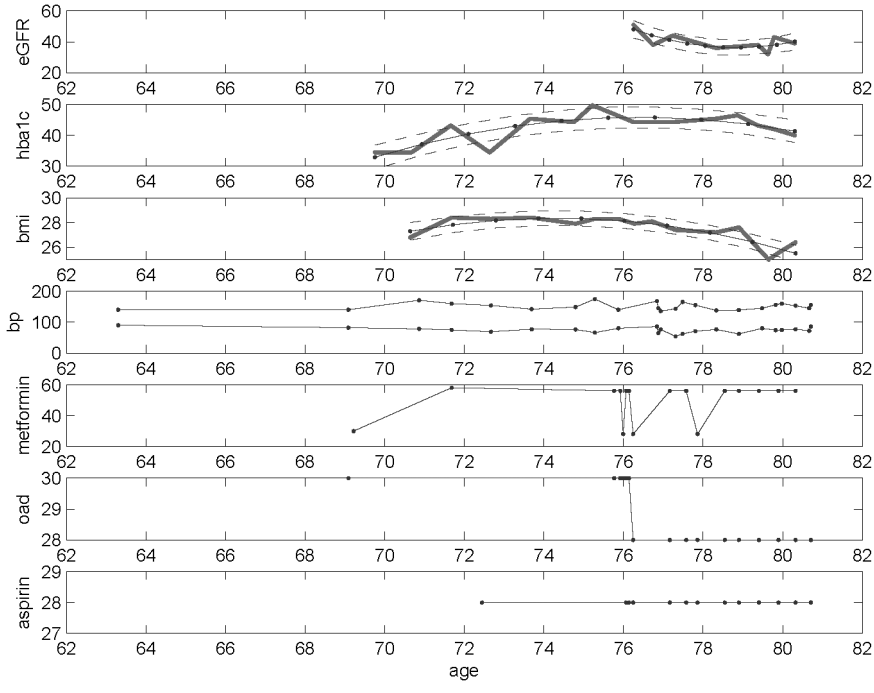

Figure 4 Pilot visualisation of health data for a person with diabetes and chronic kidney disease: charts to help practice personalised medicine. The chart shows the patient's age, any data for eGFR, HbA1c, BMI, BP, metformin, OADs, or aspirin

Implications of the findings

Longitudinal data is uniquely and readily available in primary care and other computer systems and can be harnessed to improve individual patient care. eGFR is a 'noisy' variable with a great deal of fluctuation. So many people appear to exceed the definition of 'rapidly declining' eGFR that we think this requires review. This early pilot suggests that if they can be shown to be valid, that trend lines and confidence intervals of the sort we have displayed may be of great value to clinicians in managing individual patients.

Comparison with the literature

We have already reported issues with creatinine fluctuation at the population level and these individual patient trends are compatible with this. Much of the important literature about the rationale for visualisation has been reviewed by Samal *et al* and this work is entirely compatible with their position paper. Personalised medicine is likely to be here to stay and this initiative may provide pointers towards how we can observe key data. Similar approaches have been proposed in cancer management.[23] In developing this intervention further we will need to follow advice and best practice on how to develop a new complex intervention.[24]

There are some similarities in this approach and those adopted in statistical process control (SPC)[25,26] though statistical process control proposes methods for observing data and does not encompass the potentially automated methods of data extraction which form part of this method.

Limitations of the method

There are many limitations to this pilot. We don't know if our assumptions about the underlying trends are correct, or whether our confidence intervals are valid. We have not optimised the list of variables we display. These data have been extracted from a research database which is completely under our control rather than involving live data extraction from a computerised medical record system. Most importantly no studies have yet been carried out, either simulated or in clinical practice, suggesting that this approach improves patient care.

Call for further research

It should be possible to develop modelling software capable of analysing and modelling NHS-extracted data with time-varying covariates and factors, as well as other contextual factors including economic deprivation and lifestyle. The software can be used by researchers in medical informatics, biomedical science and primary care doctors, giving an overview of the known history of a patient. This represents a major change considering that a typical piece of software in primary care practices provides only snapshot views of a patient divided into different tabs of measurements and prescriptions, thus failing to give an overall longitudinal picture of the patient's history compared with those generated here (Figure 4).

Such tools might be used to distinguish patients with CKD who are more likely to progress to complications and more serious illness from those who are less likely. There may also be potential to quantify the impact of known and new associations with progression.

Conclusions

This study is a small first step towards the generation of visualisation and modelling tools that can model longitudinal data on a personal basis. There is potential for this to be used to offer more personalised care. However, there are many stages needed first, to develop and test a complex intervention. eGFR is a noisy variable with considerable fluctuation, current definitions of 'rapid decline' as a threshold for referral appear incorrect. Computer generated visualisation of longitudinal data may provide a mechanism for improving care.

ACKNOWLEDGEMENTS

Dr Damian Fogarty of the Renal Registry, the QICKD trial senior investigators (Drs Tomson, Gallagher and Harris), and Kingston University colleagues (Dr McNiece *et al*) for their help and advice. The QICKD trial is principally funded by the Health Foundation.

REFERENCES

1 Hamburg MA and Collins FS. The path to personalized medicine. *New England Journal of Medicine* 2010; 363(4):301–4.
2 de Lusignan S and Chan T. The development of primary care information technology in the United Kingdom. *J Ambul Care Manage.* 2008;31(3):201–10.
3 Schade CP, Sullivan FM, de Lusignan S and Madeley J. e-Prescribing, efficiency, quality: lessons from the computerization of UK family practice. *Journal of the American Medical Informatics Association* 2006;13(5): 470–5.

4 Gomez GB, de Lusignan S and Gallagher H. Chronic kidney disease: a new priority for primary care. *British Journal of General Practice* 2006;56(533):908–10.
5 Gallagher H, de Lusignan S, Harris K and Cates C. Quality-improvement strategies for the management of hypertension in chronic kidney disease in primary care: a systematic review. *British Journal of General Practice* 2010;60(575):e258–65.
6 Glynn LG, Buckley B, Reddan D *et al.* Multimorbidity and risk among patients with established cardiovascular disease: a cohort study. *British Journal of General Practice* 2008;58(552):488–94.
7 National Collaborating Centre for Chronic Conditions. *Clinical guideline 73: Early identification and management of chronic kidney disease in adults in primary and secondary care.* London: National Institute for Health and Clinical Excellence (NICE), 2008.
8 de Lusignan S, Gallagher H, Stevens P *et al. Chronic Kidney Disease Frequently Asked Questions.* London: British Medical Association and NHS Employers, 2011, www.nhsemployers.org/SiteCollectionDocuments/Chronic_kidney_disease_FAQs%20-%20ja040711.pdf
9 Crinson I, Gallagher H, Thomas N and de Lusignan S. How ready is general practice to improve quality in chronic kidney disease? A diagnostic analysis. *British Journal of General Practice* 2010;60(575):403–9.
10 Tahir MA, Dmitrieva O, de Lusignan S *et al.* Confidence and quality in managing CKD compared with other cardiovascular diseases and diabetes mellitus: a linked study of questionnaire and routine primary care data. *BMC Family Practice* 2011;5:12–83.
11 de Lusignan S, Gallagher H, Chan T *et al.* The QICKD study protocol: a cluster randomised trial to compare quality improvement interventions to lower systolic BP in chronic kidney disease (CKD) in primary care. *Implement Sci.* 2009;14(4):39.
12 de Lusignan S, Tomson C, Harris K, van Vlymen J and Gallagher H. Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clin Pract.* 2011;117(3):c213–24.
13 Samal L, Wright A, Wong B, Linder J and Bates D. Leveraging electronic health records to support chronic disease management: the need for temporal data views. *Informatics in Primary Care* 2011;19:65–74.
14 Poh N and Kittler J. A method for estimating authentication performance over time, with applications to face biometrics. In: Rudela L, Mery D and Kittler J (eds) *CIARP 2007;LNCS(4472):330–69.* Heidelberg: Springer, 2007.
15 Poh N, Kittler J, Marcel S, Matrouf D and Bonastre J-F. Model and score adaptation for biometric systems: coping with device interoperability and changing acquisition conditions. *Proceedings of International Conference on Pattern Recognition (ICPR)* 2010:1229–32.
16 Poh N and Kittler J. Incorporating variation of model-specific score distribution in speaker verification systems. *IEEE Transactions on Audio, Speech and Language Processing (ASLP)* 2008;16(3):594–606.
17 Poh N, Bourlai T, Kittler J *et al.* Benchmarking quality-dependent and cost-sensitive multimodal biometric fusion algorithms. *IEEE Transactions on Information Forensics and Security (IFS)* 2009;4(4):849–66.
18 Poh N and Kittler J. A unified framework for biometric expert fusion incorporating quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 2011 (accepted for publication).
19 MySQL (Structured Query Language) database. www.mysql.com/
20 Emslie-Smith AM, Boyle DI, Evans JM, Sullivan F and Morris AD. DARTS/MEMO Collaboration. Contraindications to metformin therapy in patients with Type 2 diabetes – a population-based study of adherence to prescribing guidelines. *Diabet Med.* 2001;18(6):483–8.
21 de Lusignan S, van Vlymen J, Hague N and Dhoul N. Using computers to identify non-compliant people at increased risk of osteoporotic fractures in general practice: a cross-sectional study. *Osteoporos Int.* 2006;17(12):1808–14.
22 Current Controlled Trials. *Comparing Intervention to Lower Systolic Blood Pressure in Chronic Kidney Disease (CKD): a cluster randomised trial (CRT) – ISRCTN56023731.* www.controlled-trials.com/ISRCTN56023731/
23 Madhavan S, Gusev Y, Harris M *et al.* G-DOC: a systems medicine platform for personalized oncology. *Neoplasia* 2011;13(9):771–83.
24 Campbell NC, Murray E, Darbyshire J *et al.* Designing and evaluating complex interventions to improve health care. *BMJ* 2007;334(7591):455–9. doi: 10.1136/bmj.39108.379965.BE
25 Thor J, Lundberg J, Ask J *et al.* Application of statistical process control in healthcare improvement: systematic review. *Quality and Safety in Health Care* 2007;16(5):387–99.
26 Tennant R, Mohammed MA, Coleman JJ and Martin U. Monitoring patients using control charts: a systematic review. *Int J Qual Health Care.* 2007;19(4):187–94.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Department of Health Care Management and Policy
University of Surrey
Guildford
Surrey GU2 7XH
UK
Tel: +44 (0)1483 68 60 42
Fax: +44 (0)1483 68 60 31
Email: s.lusignan@surrey.ac.uk

Accepted November 2011

Colour versions of Figures 3 and 4 can be found online at: www.radcliffepublishing.com/journals/J12_Informatics_in_Primary_Care/Supplementary%20Papers.htm