A decision support tool for health service re-design

Demir, E., Chahed, S., Chaussalet, T.J., Toffa, S.E. and Fouladinajed, F.

This is an author's accepted manuscript of an article published in Journal of Medical Systems, 36 (2), pp. 621-630, 2012.

The final publication is available at Springer via:

https://dx.doi.org/10.1007/s10916-010-9526-8

A decision support tool for Health Service Re-design

Eren Demir, Salma Chahed, Thierry Chaussalet and Sam Toffa
Department of Information Systems and Computing, University of Westminster,
115 New Cavendish Street, London, W1W 6UW
Email: {demirer,chaheds,chausst,s.toffa}@wmin.ac.uk

Farid Fouladinajed
NHS Hounslow, 15 -21 Staines Road, Hounslow, Middlesex, TW3 3HR
Email: farid.fouladinejad@hounslowpct.nhs.uk

Abstract

Many of the outpatient services are currently only available in hospitals, however there are plans to provide some of these services alongside with General Practitioners. Consequently, General Practitioners could soon be based at polyclinics. These changes have caused a number of concerns to Hounslow Primary Care Trust (PCT). For example, which of the outpatient services are to be shifted from the hospital to the polyclinic? What are the current and expected future demands for these services? To tackle some of these concerns, the first phase of this project explores the set of specialties that are frequently visited in a sequence (using sequential association rules). The second phase develops an Excel based spreadsheet tool to compute the current and expected future demands for the selected specialties. From the sequential association rule algorithm, endocrinology and ophthalmology were found to be highly associated (i.e. frequently visited in a sequence), which means that these two specialties could easily be shifted from the hospital environment to the polyclinic. We illustrated the Excel based spreadsheet tool for endocrinology and ophthalmology, however, the model is generic enough to cope with other specialties, provided that the data are available.

Keywords: Service re-design, sequential association rule, demand and capacity, decision support tool

1 Introduction

Primary care is the care provided by people we normally see when we first have a health problem. It may just be a visit to your local doctor, dentist, or an optician for an eye test. In England, all of these services are managed and commissioned by local primary care trusts (PCTs) on behalf of the National Health Service (NHS). There are currently 152 PCTs in England in control of 80% of the NHS budget (approximately £83.2bn; $137bn) [1]. PCTs work in close collaboration with local authorities and other agencies to provide health and social care services (e.g. continuing care services). As they are local organisations, they are best positioned to understand the needs of their community, so they can make sure that the organisations providing health and social services are working effectively. For instance, they must make sure that service capacity meets people needs within their area and that these services are accessible.

The recession and the banking crisis forced the UK government to cut funding on public services. During 2010-11, the NHS will need to contribute £2.3bn ($3.8bn) of the £5bn of public sector efficiency savings [2], where the highest savings are expected primarily from PCTs. In anticipation for tough times ahead, it is in the interest of PCTs to obtain evidence based knowledge of the use of their services (e.g. accident & emergency, inpatients, outpatients, etc.) based on regions and patient groups in order to reduce the inequalities in health outcomes, improve matching of supply and demand [3], and most importantly reduce costs generated by its various services. These long term objectives are related to the vision of World Class Commissioning (WCC) [4].

To address some of these issues, the Department of Health in England plans to re-design the current delivery of primary care through the introduction of "polyclinics". Currently, general practice (GP) doctors deliver primary care services by providing treatment and drug prescriptions and where necessary patients are referred to specialists, such as for outpatient care, which is provided by local hospitals (or specialised clinics). However, general practitioners (GPs) are limited in terms of size, resources, and the availability of the complete spectrum of care within the local community. With the re-design of primary care services, GPs will be based at polyclinics, where the range of services available are expected to exceed that of most existing GP practices. Polyclinics aim to offer access to antenatal and postnatal care, healthy living information and services, community mental health services, community care, social care and specialist advice all in one place. They should provide the infrastructure (such as diagnostics and consulting rooms for outpatients) to allow a shift of services out of hospital settings [5].

In April 2009, Hounslow PCT introduced one of London's first polyclinics that began providing a wide range of primary care services. It is responsible for managing and commissioning health care services for around 224,546 people [6] with a diverse range of ethnic minorities (50% of its population), living in eight districts (i.e. Bedfont, Brentford, Chiswick, Feltham, Hanworth, Heston, Hounslow and Isleworth). Chiswick, in the east, is a mainly prosperous suburb, whereas Central Hounslow, Heston and Cranford are ethnically very diverse, and the most deprived.

Many of the outpatient services offered by Hounslow PCT are only available in hospitals. Some of these services are expected to shift out of their hospital setting to the newly opened polyclinic. However, a concern to the PCT is which of the outpatient services are to be shifted. The plan is to provide services that are highly interrelated, i.e. regularly visited in a sequence. For instance, a diabetic patient may need to consult an endocrinologist and an ophthalmologist the same day to avoid visiting the polyclinic several times in a week. This will inevitably reduce cost and waiting lists, and most importantly improve patient experience. Furthermore, it is in the interest of the PCT to quantify the current (short term) and future (long term) demands for the shifted services. Hence, understanding the short and long term consequences of shifting services will enable the PCT to allocate resources (doctors, nurses, equipment, time slots, etc.) accordingly. This paper will focus on two main areas. The first is to extract unknown patterns of the use of acute activities with the intensity of usage based on patient characteristics (e.g. age), and the patterns of uses of care, in particular patient pathways, i.e. the sequence of visited specialties. The second is to determine the current and expected future demands for the sequence of visited specialties.

As part of this work, the complete registered population of Hounslow data is used. Since we are dealing with a large amount of data, data mining methods can be a useful approach to discover hidden patterns and rules to support the PCT for decision making purposes. Data mining methods have been applied to a variety of domains (e.g. direct mail, retail, insurance) including healthcare. It is the analysis of (often large) observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [7]. There are five main principles in the process of data mining: business understanding (e.g. objectives); data understanding; data preparation; modelling and evaluation. For instance, given the registered population of Hounslow PCT (224,546 individuals), such as age, gender, socioeconomic status (SES), employment status, location of residence, etc., and we may be interested to extract the unknown patterns from this large dataset. For example, using the five principles of data mining process, and applying cluster analysis, we may observe that there are two distinct groups, e.g. West Hounslow may be highly dominated by younger people, semi-skilled professionals, low in SES, whereas East Hounslow may comprise more middle-aged people, professional workers and high in SES. This can be considered to be a valuable piece of information to better understand the diversity of a region with respect to its population.

With respect to healthcare applications, [8] and [9] presented a selection of data mining techniques that could be suitable for analysing healthcare data. In another case, clustering methods were used to detect similarities of community centres of a Slovenian region in terms of availability and accessibility of public health care resources [10]. Then a decision tree learning algorithm was used to explain the main differences between those clusters. Note that clustering methods partition observations into groups (called clusters) to maximise observations' similarity within clusters and dissimilarity between clusters, while a decision tree learning algorithm classifies the target variable based on the partitioning of the input variables into a set of homogenous regions (a tree structure).

Association rules were used by [11] to examine and detect incorrect or fraudulent billing patterns within a particular specialist group, in order to reduce inappropriate payments by the (Australian) Medicare and to clarify specific areas of the Medicare Benefits Schedule. Another example is the identification of frequent associations between diagnoses of mothers and their newborns. The aim was to determine the risk factors leading to mother and/or newborn health complications [12].

A minor modification of the association rule algorithm has resulted in the development of sequential association and inter-transactional association rules, which explores time-dependent patterns. [13] has introduced the approach for mining sequential patterns over large databases. Here, the transactions of a customer were ordered by increasing transaction-time and assumed to be a sequence. Each transaction is a set of items purchased together at the same time. The problem is to capture the most frequent sequences supported by the majority of customers. In another study, sequential association rules were used to determine the aircraft failure patterns, which could lead to the prediction of failures of aircrafts under various scenarios (e.g. aircraft mission, season) [14]. This study was aimed to support future demand forecasting for aircraft spare parts. In healthcare research, [15] presented a framework (combining sequential association rules and classification trees) for predicting patient pathways between different medical units. In our case, the aim is to extract the set of specialties that are frequently visited in a sequence, hence sequential association rules is an adequate algorithm to use.

The next section describes the association rule algorithm; the demand and capacity modelling approach, and the Hounslow dataset; Section 4 illustrates the result, where the sequence of frequently visited specialties are extracted in section 4.1 and the spreadsheet model for the decision support tool is illustrated in sections 4.2 and 4.3, and finally discussion in section 5.

3. Materials and Methods

There are two phases to this study:

- Phase 1: explores the set of specialties that are frequently visited in a sequence. For this purpose, a sequential association rule algorithm is adopted (unsupervised data mining algorithm), which is summarised in section 3.1.
- Phase 2: develops a simple user friendly decision support tool to compute the current demand, the expected future demand, and growth projections (i.e. how many more attendances could be expected one year later) for the selected specialties discovered in Phase 1. This is based on regression analysis, confidence intervals and scenario planning, which are explained in section 3.2.

3.1. Modelling the association between specialties

Association rules are a simple but effective unsupervised data mining technique, which has the ability to discover associations between items that are stored in the database. For example, a supermarket manager may want to know if certain groups of items are consistently purchased together, where this information can be used to inform store layouts (cross-selling). A special case of association rules is market basket analysis, which is often applied to very large commercial datasets with binary value variables (i.e. 0 or 1).

Suppose we have the binary variables $Z_1,...,Z_K$ (e.g. each variable can be considered to be an item and 0 and 1's would indicate whether the item is in the basket or not). Formally, the aim of finding modes in the database becomes finding subsets $M \subseteq \{1,...,K\}$, such that $P\left[\bigcap_{m\in M}(Z_m = 1)\right]$ is large, which is estimated by $S(M) = \frac{\#\left\{\bigcap_{m\in M}(Z_m = 1)\right\}}{N}$, where N is the number of records in the dataset, i.e. the fraction of observations for $\left\{\bigcap_{m\in M}(Z_m = 1)\right\}$ is true, and $S(M)$ is called the support of the itemset M. Therefore, the aim is to find all itemsets that have large support in the database. The items Z_m $(m \in M)$ are partitioned into two disjoint subsets A and B (that is, $A \cup B = M$) and is written as $A \Rightarrow B$, where A is called the *antecedent* and B is called the *consequent*. For example, suppose we have an itemset containing jam (j), bread (b) and butter (t), i.e. $M = \{j,b,t\}$, and some examples of association rules could be $\{j,t\} \Rightarrow \{b\}$, $\{t,b\} \Rightarrow \{j\}$, $\{t\} \Rightarrow \{j,b\}$. The evaluation of a rule is usually based on the *support* and *confidence* of a rule. Support of a rule is $S(A \Rightarrow B) = S(M)$, i.e. how frequent the rule is in the database, and the confidence of a rule is $C(A \Rightarrow B) = \frac{S(A \Rightarrow B)}{S(A)}$, i.e. the proportion of records where B is observed among all records where A is observed.

3.2 Modelling demand and capacity

A patient may encounter multiple attendances to a clinic before being referred to a different clinic. For instance, a patient may have had k visits to the Endocrinology clinic before the first attendance to the ophthalmology clinic (j_1) (refer to Figure 1). The initial $k-1$ sequential visits to the same clinic are referred to as independent treated demand. The treated demand is the number of patients that have attended the outpatient clinic within a specified time window, e.g. 1 week. However, when the same patient moves from i_k (k^{th} Endocrinology visit) to j_1 (1st ophthalmology visit), this is referred as a dependent treated demand. Hence, the current demand and the expected future demands are provided separately for both cases. First, the independent treated demands for both endocrinology and ophthalmology are illustrated. Secondly, the dependencies between the two specialties are considered.

*******Figure 1 goes here*******

There are many cases where patients fail to attend their appointment, which means that resources allocated to that patient are not used. To avoid any waste and for better management of resources, we need to determine the expected future demand for did not attend (DNA) cases, as well as for those attended cases. There are 7 categories of attended or did not attend codes in the outpatient dataset. These are:

- 5 = Attended on time or, if late, before the relevant care professional was ready to see the patient
- 6 = Arrived late, after the relevant care professional was ready to see the patient, but was seen
- 7 = Patient arrived late and could not be seen
- 2 = Appointment cancelled by, or on behalf of, the patient
- 3 = Did not attend - no advance warning given
- 4 = Appointment cancelled or postponed by the Health Care Provider
- 0 = Not applicable - Appointment occurs in the future.

Category 5 and 6 were considered as attended; 7, 2, and 3 were considered as did not attend, and category 4 was not taken into account as the cancellation of the appointment was not due to the patient.

Therefore, the current demand and the expected future demand are provided for each of the following:

- Independent treated demand
- DNA related to independent treated demand
- Dependent treated demand
- DNA related to dependent treated demand.

The current demand is estimated using confidence intervals in the weekly numbers of patient visiting the outpatient clinics. The expected future demand is based on linear regression, assuming that there is a linear increase in the demand against time (weeks).

3.3 Data

Hounslow PCT database (namely Hounslow Atlas Database) comprises eight sources of data: inpatient care, outpatient care, accident & emergency (A&E), GP details, registered population, community, mental health in hospital, and mental health in the community. The data was provided in Microsoft Access and Excel format and necessary steps were taken to import the data into MySQL version 5.0, so that database programming could easily be carried to prepare the data for analysis. The intention of Hounslow PCT is to shift services from outpatient care to the polyclinic. Hence, the outpatient and the registered population datasets will be used. The data period is from 01/04/06 to 31/08/08. The total number of attendances during the specified data period is 934,536. Although a patient residing outside the boundaries of Hounslow could also make use of outpatient services, Hounslow PCT is only interested in its own population. Thus, when the outpatient dataset is linked with the registered population, 934,536 reduce to 829,832.

3.4 Data preparation

There are a number of criteria for the data preparation process. First, we extract all those patients that have experienced at least 2 different specialties, so that the associated specialties can be identified through the association rule algorithm. Second, we are only interested in the sequence of outpatient attendances. Table 1 illustrates the sequence of attendances of a patient. The first two attendances are specialty code 110 (trauma & orthopaedics). Since the interest lies in the sequence of different specialties, and not within the same specialty, these two attendances are joined as a single record. Hence, Table 1 collapses to Table 2. Each record in Table 2 is either based on joining a number of same specialties that were visited in a sequence (110,110 □ 110). Hence, based on these criteria's 829,832 attendances reduce to 364,767 records. Furthermore, due to missing NHS Numbers (29,251) and treatment function codes (8,177), these reduce to 327,339 records.

Table 1 Sequenced attendances of a patient. TFC refers to treatment function code (specialty code)

NHS Number	TFC	Attendance date
2050617821	110	2007-07-11
2050617821	110	2007-08-01
2050617821	150	2007-08-14
2050617821	130	2007-09-05
2050617821	307	2007-09-20
2050617821	307	2007-10-23
2050617821	130	2007-10-26
2050617821	307	2007-12-04
2050617821	307	2008-02-25

Table 2 Collapsed data

NHS Number	TFC	Attendance date
2050617821	110	2007-08-01
2050617821	150	2007-08-14
2050617821	130	2007-09-05
2050617821	307	2007-10-23
2050617821	130	2007-10-26
2050617821	307	2008-02-25

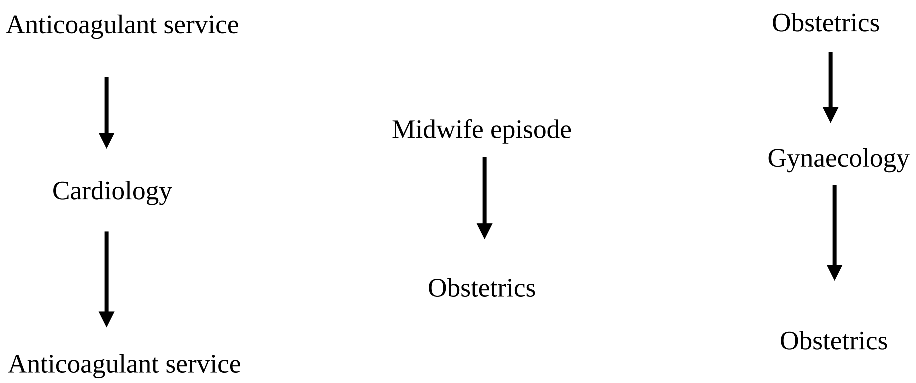
4 Results

The results are illustrated in three subsections, the associated specialties, data related to the dependent and independent demand for the associated specialties, and the capacity model (i.e. the Excel spreadsheet tool).

4.1 The associated specialties

This section illustrates the obvious associated specialties, i.e. those specialties that would be expected to have a relationship (validation purposes) and the not so obvious relations, i.e. the unknown relationships between specialties. These relationships are based on expert's judgements.

4.1.1 The obvious relations



Support = 1.34%, 357 instances Confidence = 90.6 %	Support = 1.89%, 394 instances, Confidence = 71.2 %	Support = 1.94%, 325 instances, Confidence = 57 %

As mentioned earlier the strength of the association is based on the *support* and *confidence* of the rule. The *support* is defined as the proportion of records in the data set which contain the rule. For example, the itemset Z = {Anticoagulant service, Cardiology, Anticoagulant service}, which is partitioned into two disjoint subsets with the antecedent $A =$ {Anticoagulant service, Cardiology}, consequent $B =$ {Anticoagulant service}, and support $S(A \Rightarrow B) = 0.0134$, since it occurs in 1.34% of all the records (1.34 out of 100 records), which is 357 instances. We should bear in mind that there are very large numbers of sequence of attendances (possibly thousands) and a support of 1.34% can be considered to be high.

The *Confidence* of a rule is the proportion of records where {Anticoagulant service, Cardiology, Anticoagulant service} is observed among in all records where {Anticoagulant service, Cardiology} is observed. Therefore, when a patient visits the anticoagulant service and cardiology together (in a sequence), 90.6% of the time patients visit again the anticoagulant service. Table 3 illustrates a number of other associated specialties. We should note that only a small number of associated specialties are illustrated (confidence greater or equal to 50%), where these relationships were carefully analysed with domain experts.

Table 3 Obvious associated specialties. TO refers to trauma & orthopaedics

Antecedent (A)	**Consequent (B)**	**Support (%)**	**Confidence (%)**
{TO, Ophthalmology, TO}	{Ophthalmology}	1.79	62.2
{Ophthalmology, Rheumatology}	{Ophthalmology}	1.65	62.2
{Ear nose and throat, Paediatrics}	{Ear nose and throat}	2.03	56.6
{General medicine, Ophthalmology}	{Ophthalmology}	2.33	53
{Ophthalmology, Gastroenterology}	{Ophthalmology}	2.08	51.7
{Orthoptics}	{ Ophthalmology}	1.97	71.1

4.1.2 Not so obvious relations

Ophthalmology was linked with a variety of specialties (e.g. endocrinology, urology, cardiology, ENT, etc.) (see Table 3). The highest confidence is 67.4%. When a patient visits the endocrinology clinic and ophthalmology clinic together (in a sequence), 67.4% of the time patients visit again the endocrinology clinic.

Endocrinology → Ophthalmology → Endocrinology
Support = 1.98%,
357 instances
Confidence = 67.4%

Urology → Ophthalmology → Urology
Support = 2.04%,
358 instances
Confidence = 60 %

Cardiology → Ophthalmology → Cardiology
Support = 2.5%,
374 instances
Confidence = 51%

Supported by expert opinion, this provided evidence that endocrinology and ophthalmology were among the specialties that could be shifted from the hospital environment to the newly opened polyclinic. The justification is that it is highly associated (confidence 67.4%), both are age related conditions (homogenous group of patients), these clinical conditions do not require expensive resources (e.g. equipment), and most importantly endocrinology and ophthalmology could easily be handled outside the hospital setting, simply because these patients do not usually experience complications. Hence, our modelling approach for estimating the current and future demand will focus on these two specialties.

4.2 Independent and dependent demand

The number of attended and DNA cases for endocrinology and ophthalmology for each week, based on various age groups (0-34, 35-49, 50-59, 60-69 and 70+**)** was extracted. We have tested with various other age groups and observed that the chosen age groups were the most illustrative for the independent cases, i.e. similar number of weekly attendances and DNA for each age group. However, these age groups were not representative for the dependent cases. There were very few weekly dependent attendances for the age group 0-34 (many 0 attendances per week), and so the age groups 0-34 and 35-49 were combined. Table 4 illustrates a summary of attendances and DNA for independent cases. The total number of attended cases for endocrinology and ophthalmology are 10,184 and 51,022, respectively, whereas 3,977 and 14,183 for DNA, respectively.

Table 4. Independent treated demand of endocrinology and ophthalmology clinic (Attended and DNA)

	0-34	35-49	50-59	60-69	70+	Total
Total number of independent endocrinology attendances	1894	2737	2054	1943	1556	10184
Total number of independent endocrinology DNA	841	1131	782	687	536	3977
Total number of independent ophthalmology attendances	7785	5246	6429	9920	21642	51022
Total number of independent ophthalmology DNA	2978	1918	1961	2381	4945	14183

The number of attended and DNA cases for those patients moving from endocrinology to ophthalmology (and vice-versa) for each week, per age group (0-49, 50-59, 60-69, and 70+) was extracted. We do not assume a gap (time) between the visits. The rationale is that we are only considering 2.5 financial years, and an arbitrarily chosen time window to define a gap could lead to the exclusion of a large number of cases, which may result in an inaccurate estimation of demand. Suppose a patient attended the endocrinology clinic on 1st March 08 and attended the ophthalmology clinic the following month (13th

April 08). Although there is a gap of 6 weeks, we count this as an attendance for week 3 of April 08. From Table 5, we observe 2,550 sequential movements from endocrinology to ophthalmology, and 917 DNA cases.

Table 5. Dependent treated and not treated demand for each of the age groups for the association endocrinology – ophthalmology

	0-49	50-59	60-69	70+	Total
Total number of dependent endocrinology ↔ Ophthalmology attendances	508	498	781	763	2550
Total number of dependent Endocrinology ↔ Ophthalmology DNA	253	210	220	234	917

4.3 The decision support tool

The model was implemented within the Excel spreadsheet environment. The first three worksheets are related to the estimation of the demand, namely “Demand”, “Expected Future Demand” and “Scenario Based Future Demand”. The remaining six are the data extracts. The first worksheet named “Demand” presents the estimated weekly demands for

- Independent treated demand of the endocrinology clinic
- Independent treated demand of the ophthalmology clinic
- DNA related to independent treated demand of the ophthalmology clinic
- DNA related to independent treated demand of the endocrinology clinic
- Dependent treated demand of the endocrinology and ophthalmology clinics
- DNA related to dependent treated demand of the endocrinology and ophthalmology clinics.

The second worksheet, “Expected Future Demand”, gives an estimate of the expected future demand (attended/DNA) for both the independent and dependent cases. It is based on linear regression, assuming that there is a linear increase in the demand against time (weeks). The third worksheet, “Scenario Based Future Demand”, gives an estimate on the expected future demand (attended/DNA) for both the independent and dependent cases based on the user definition of weekly increase, hence a scenario based tool. The spreadsheet model allows the user to specify a date range (From - To), so that the current demand can be estimated either using the full dataset (April 06 – August 08) or a specific range. Expected future demand worksheet uses the full dataset to estimate (linear regression) growth parameters. The user specifies the number of weeks to forecast (e.g. 24 weeks) and the growth parameters are used to output the expected future demands. "Scenario based future demand" worksheet produces similar results as the expected future demand worksheet, however the growth parameters are user defined.

4.3.1 Current demand

Weekly numbers of patients visiting the outpatient clinics for the independent and dependent (attended and DNA) cases were assumed to be normally distributed. A visual depiction of the data (histogram) and q-q plots (quantile to quantile) were used to verify this assumption. The mean in this context refers to the average weekly attendances (or DNA cases). However, instead of just estimating the average weekly attendances, we may want to examine the reliability of the estimated weekly average. Thus, confidence intervals (CI) are used to indicate the reliability of the estimated parameter. The spreadsheet model allows the user to specify a CI limit. As the treated demand does not possess an increasing trend or a seasonal pattern, weekly averages and CI provide a relatively accurate estimate.

The worksheet "Demand" presents the current demands for both the independent and dependent cases. The outputs on the left of Figure 2 are for the independent treated demand (attended/DNA cases) for endocrinology (top) and ophthalmology (bottom). The outputs on the top right of Figure 2 are for the dependent treated demand (attended/DNA cases). For instance, using the past 24 weeks (March 08 – August 08) historical attendances to endocrinology for the age groups 0-34, 35-49, 50-59, 60-69, and 70+, the estimated weekly average treated demands are 18, 27, 21, 19, and 14 attendances, respectively. The Limit specifier = 0.025 estimates the 95% confidence interval, with a CI [18, 24] for the age group 50-59. Therefore, we are 95% certain that the weekly attendance to the endocrinology clinic is between 18 and 24 attendances per week for the age group 50-59. In the case of DNA, Hounslow PCT could be expecting on average 9, 11, 7, 6, and 4 DNA cases to endocrinology per week for the age groups 0-34, 35-49, 50-59, 60-69, and 70+, respectively. Hounslow PCT could also expect between 6 to 8 (95% CI) patients to visit both the endocrinology and ophthalmology clinics (or vice versa) in a given week for the age group 60-

69. The values in the middle of each section (e.g. Attended) is the average, whereas the top and bottom are the lower and upper limits (confidence intervals), respectively.

*******Figure 2 goes here*******

4.3.2 Expected Future Demand

The "Expected Future Demand" worksheet provides an estimate in the expected future growth on demand. The methodology is based on linear regression using the past history of weekly attendances/DNA. Here, the full dataset is utilised to estimate the growth parameters. The user could specify any projection period, e.g. 24 weeks, 52 weeks, etc. For instance, using a 24 week projection period (Figure 3), Hounslow PCT could expect 20 attendances per week to the endocrinology clinic for the age group 0-34.

*******Figure 3 goes here*******

4.3.3 Scenario based future demand

The scenario based future demand worksheet produces similar results to the expected future demand worksheet. However, the growth parameters are user defined. The top half of the worksheet (Figure 4) is for the user to input the growth parameters, where these parameters are used to compute scenario based future demand forecasts (bottom half). For instance, you may expect no weekly increase (growth rate=0) for all age groups (both independent and dependent cases) apart from the age group 0-34 for endocrinology clinic where an increase of 1 attendance per week is expected. Based on this scenario, with a projection period of 24 weeks, endocrinology clinic could expect 42 attendances (6 months later).

*******Figure 4 goes here*******

5 Discussion

Most of the outpatient services provided by Hounslow PCT as well as other PCTs are only available in hospitals. However, there are plans to provide some of these services alongside with General Practitioners within polyclinics. Over the long term, these changes would reduce the inequalities in health outcomes, improve the matching of supply and demand and reduce costs. This changing context has caused some concerns to Hounslow PCT, particularly to decide which of the outpatient services are to be shifted from hospital setting to the polyclinic and their associated current and expected demands. The objective of this project was to assist Hounslow PCT in their service re-design decision making process. In this paper, we have described the approach to the solution and the related decision support tool developed.

From Phase 1 of the project, we have observed that endocrinology and ophthalmology were found to be highly associated specialties, where patients were frequently visiting these two clinics. One of the main advantages of the association rule is the ability to quantify the strength of an association in a simple and concise manner using the support and confidence of a rule, where these quantities could easily be understood by non-technical audience, e.g. the support is defined as the proportion of records in the data set which contain the rule. However, the downside of the algorithm is that it requires vast amount of data to compute the associated specialties. During the second phase, we have developed an Excel based spreadsheet model that could facilitate the computation of current demand and the projection of expected future demands for the two specialties (independently and dependently). Using historical data and the assumption of a normally distributed weekly attended/DNA cases, we developed a spreadsheet model that allows the user to specify a date range, which would then be used for the estimation of model parameters. As the treated demand does not possess an increasing trend or a seasonal pattern, the weekly average and a CI provided relatively accurate estimates. The current demands and growth projections were provided for various age groups.

In this paper, only two specialties, endocrinology and ophthalmology were considered, however the model is generic enough to cope with other specialties, provided that the data is available. The Excel based spreadsheet model is easy to manipulate and does not require a huge amount of input data. However, there are two potential problems or concerns that could arise as a result of making a decision using this tool; 1) the user will need to ensure that sufficient data are available to estimate the expected demand, so that the model computes reliable and accurate forecasts, else decisions can be made based upon inaccurate estimates, and 2) just like any other decision support tool, inappropriate interpretations of the estimates can be a serious concern, particularly if the user is not familiar with the statistical terminologies, e.g. association rules, 95% confidence interval, regression analysis.

In the medium term, this project will serve as a basis for scheduling decision support tool. In fact, the care needs of the various age groups regarding a particular specialty could differ slightly, hence by assigning different time slots to each age group for every specialty it could improve the quality of care service. PCT managers will be able to allocate these time slots to specialities by evaluating the weekly treated demand per age group and by considering the association between specialties. For instance, endocrinology and ophthalmology clinic could be scheduled for the same day to treat patients aged 0-69 years, who require care related to both specialities the same week, provided that the level of demand is greater that a specific threshold (justifying the opening of these two clinics). Further work on scheduling health centres will be carried on in the future.

References

[1] *National Health Service England*, http://www.nhs.uk/NHSEngland/aboutnhs/Pages/Authoritiesandtrusts.aspx (accessed 18 August 2009).

[2] C. Ham, The 2009 budget and the NHS, British Medical Journal 338 (2009) 1024-1025.

[3] JM Nguyen, P. Six, D. Antoniolib, P. Glemain, G. Potel, P. Lombrail and P. Le Beux, A simple method to optimize hospital bed capacity, International Journal of Medical Informatics 74 (2005) 39-49.

[4] *Department of Health, World Class Commissioning*, http://www.dh.gov.uk/en/Managingyourorganisation/Commissioning/Worldclasscommissioning/DH_083204, 2009.

[5] NHS Evidence, http://www.library.nhs.uk/HealthManagement/ViewResource.aspx?resID=267332 (accessed 18 August 2009).

[6] NHS Hounslow, Annual Public Health Report 2008/2009, http://www.hounslowpct.nhs.uk/hpct_public_health.html, 2009.

[7] DJ Hand, H. Mannila and P. Smyth. (2001) Principles of Data Mining, pg 1, MIT Press.

[8] A. Lavrac, Selected techniques for data mining in medicine, Artificial Intelligence in Medicine 16 (1999) 3-23.

[9] MK. Obenshain, Application of data mining techniques to healthcare data, Infection Control and Hospital Epidemiology 25 (2004) 690-695.

[10] N. Lavrac, M. Bohanec, A. Pur, B. Cestnik, M. Debeljak and A. Kobler, Data mining and visualization for decision support and modelling of public health-care resources, Journal of Biomedical Informatics 40 (2007) 438-447.

[11] Y. Shan, D. Jeacocke, DW Murray and A. Sutinen, Mining medical specialist billing patterns for health service management, In Proceedings of Seventh Australasia Data Mining Conference (AusDM 2008), Glenelg, South Australia. CRPIT, **87**. Roddick, J. F., Li, J., Christen, P. and Kennedy, P. J., Eds. ACS. 105-110.

[12] R. Dogaru, D. Zaharie, D. Lungeanu, E. Bernad and M. Bari, A framework for mining association rules in data on perinatal care, The 8th International Conference on Technical Informatics (CONTI'2008), Timisoara, Romania.

[13] R. Agrawal and R. Srikant, Mining sequential patterns, In proceedings of the 11th International Conference on Data Engineering, Taipei, Taiwan, March 1995 (ICDE'95).

[14] HK Han, HS Kim and SY Sohn, Sequential association rules for forecasting failure patterns of aircrafts in Korean airforce, Expert Systems with Applications 36 (2009) 1129-1133.

[15] D. Thierry, C. Yongmei, C. Gilles and D. Patrice, Analysis of hospitalised patient flows using data-mining, Studies in health technology and informatics 95 (2003) 263-268.