An Application of KEFIR to the Analysis of Healthcare Information

Christopher J. Matheus Gregory Piatetsky-Shapiro
Dwight McNeill*

GTE Laboratories Incorporated
40 Sylvan Rd., Waltham MA 02254
matheus@gte.com, gps@gte.com, dnm0@gte.com

Abstract

The Key Findings Reporter (KEFIR) is a system for discovering and explaining "key findings" in large, relational databases. This paper describes an application of KEFIR to the analysis of health-care information. The system performs an automatic analysis of data along multiple dimensions to determine the most interesting deviations of specific quantitative measures relative to norms and previous values. It explains key findings through their relationship to other findings in the data, and, where possible, generates simple recommendations for correcting detected problems. A final written report, complete with business graphics, is produced for viewing remotely over the internet with Mosaic, or for printing to hardcopy.

Keywords: knowledge discovery, databases, health care

1 Introduction

Knowledge discovery techniques are being used successfully today to analyze and explore large databases in numerous scientific, financial, and manufacturing domains [Piatetsky-Shapiro, 1993, Matheus *et al.*, 1993, Piatetsky-Shapiro and Frawley, 1991]. In this paper we describe our recent work on applying knowledge discovery to the analysis of health-care data in a system called Health-KEFIR. Health-KEFIR was built using KEFIR, our discovery system shell for detecting, evaluating, and explaining interesting deviations in large, relational databases.

*Dwight McNeill is the Health Care Information Manager for GTE Service Corporation.

2 The Problem: Rising Health-Care Costs

Health-care costs in the US have been rising at three times the rate of inflation over the past 10 years. This has weakened US competitiveness in the global market because of the relatively higher cost of health care as compared to other countries.[1] Health-care costs now represent about 50% of corporate net profits. For our company, GTE, health-care costs in 1994 will be approximately $700 million dollars, or about $5,000 per employee. In efforts to control this escalating problem, large employers have invested millions of dollars in information systems for recording and reporting on health-care costs. In turn, medical information companies have sprung up to service this need and to provide expert health-care analysis.

Expert analysis of health-care data is time consuming and very expensive. A single report may take weeks or months to prepare and can cost tens of thousands of dollars. For large corporations, which typically order many reports for different business units, health-care consulting costs may run into millions of dollars per year. The great time and expense of preparing a report acts as a disincentive to ordering them in many cases, thus eliminating potential savings opportunities. Even when a report is ordered, it may be incomplete because an exhaustive search of possible findings and their explanations is simply infeasible by manual means. Later we will give an example where human experts overlooked an important finding detected by `Health-KEFIR`.

3 Health-Care Management Data Analysis

Current approaches to health-care data analysis rely on a set of relatively standard *measures* or *indicators* such as `Average_hospital_payments_per_capita`, `Admission_rate_per_1000_people`, and `Cesarean_section_rate` [McNeill, 1993]. These measures assess various aspects of health care, including cost, price, usage, and quality. Measures are often related by formulas to other measures, for example, `Admission_rate_per_1000 = Admissions * 1000 / Number_of_covered`. National, regional, and other norms are routinely compiled for many of these measures to serve as references for judging quality and performance.

Measures are typically aggregate values taken over populations of individuals. For a corporation, the primary population of interest is its employees and their dependents, i.e. those individuals for which the company provides health-care coverage. Various sub-populations of this group are also of interest to the company, such as separate business units, national regions, union vs. non-union employees, etc. From the health-care side, sub-populations of interest are defined in terms of standard categories, such as Inpatient/Outpatient, Inpatient Admission Type (medical, surgical, etc.), Major Diagnostic Category (MDC), and Diagnostic Related Group (DRG).

A fundamental question in health-care analysis is: For a given population, how do the standard measures compare to previous values and to normative or expected values? If a measure for the population has changed dramatically or deviates significantly from the norm, then this represents a potentially interesting *finding*. Additional factors in determining how interesting a finding is include its impact on the bottom line (i.e. how much it costs

[1] In the US, health-care costs consume a larger share of the manufacturing cost of a car than does steel.

the company in dollars), the significance of the finding (e.g. is it due to chance?), and whether there are potential intervention strategies. This last factor is particularly important because it identifies where a health-care manager can achieve real improvements, i.e. reduce cost and/or improve quality. For example, a 10% increase in costs for normal pregnancies would be less interesting than a 10% increase in costs for problem newborns, since well-known intervention methods exist for early prenatal care. The interestingness of deviations is examined in detail in [Piatetsky-Shapiro and Matheus, 1994].

In addition to uncovering the significant findings, the analyst needs to explain them to the extent possible given the data and the analyst's knowledge of the health-care field. The standard procedure for explaining a high-level finding is to "drill down" into the data. In this technique, the cause of a finding is traced to either other significant deviations in smaller sub-populations, or to other measures that drive the calculation of the first finding. This process progresses in a top-down fashion, starting with the entire population at the top level and drilling down into smaller and smaller populations until no more significant events are found. The key findings and their explanations are then compiled into a summary report along with recommendations for courses of action.

The task of deviation detection and knowledge-based drill-down is well suited for automation. A similar task is performed by Spotlight [Anand and Kahn, 1992] and CoverStory [Schmitz *et al.*, 1990], two products for identifying and reporting on trends and exceptional events in supermarket sales databases. The goal of our work is to apply some of the same techniques in the much richer context of health care in order to identify ways of reducing GTE's health-care costs and improving the quality of care to its employees.

4 The KEFIR System

KEFIR is a domain-independent system for discovering and explaining key findings in large, relational databases. Its design models the analytic process employed by the health-care analysts we have consulted with. The driving premise of the system is that many of the most interesting patterns to be found in health-care databases can be described as *deviations*. A deviation, in our use of the term, is a difference between an *observed value* and a *reference value*. The observed value is always the current value for a measure in a specified population, and it is always represented as a single numeric figure. The reference value can be of various sorts. A deviation over time occurs when the reference is from a previous quarter or year. A standard norm can serve as the reference value in a "normative deviation." Alternatively, a model could generate an expected prediction for the reference value to create a "deviation from expectation."

Deviations are powerful because they provide a simple way of identifying interesting patterns in the data. We have studied many knowledge-discovery algorithms with potential for identifying vast numbers of significant patterns from data, but most of these are unable to determine when a pattern is truly interesting to the user [Matheus *et al.*, 1993]. With deviations we have a simple way to identify things that differ from our expectations – since they differ from what we expect, they are by definition interesting at least to some degree.

A complete discovery system requires more than simple deviation detection. Typically the number of detected deviations can be quite large, and so we need a mechanism for

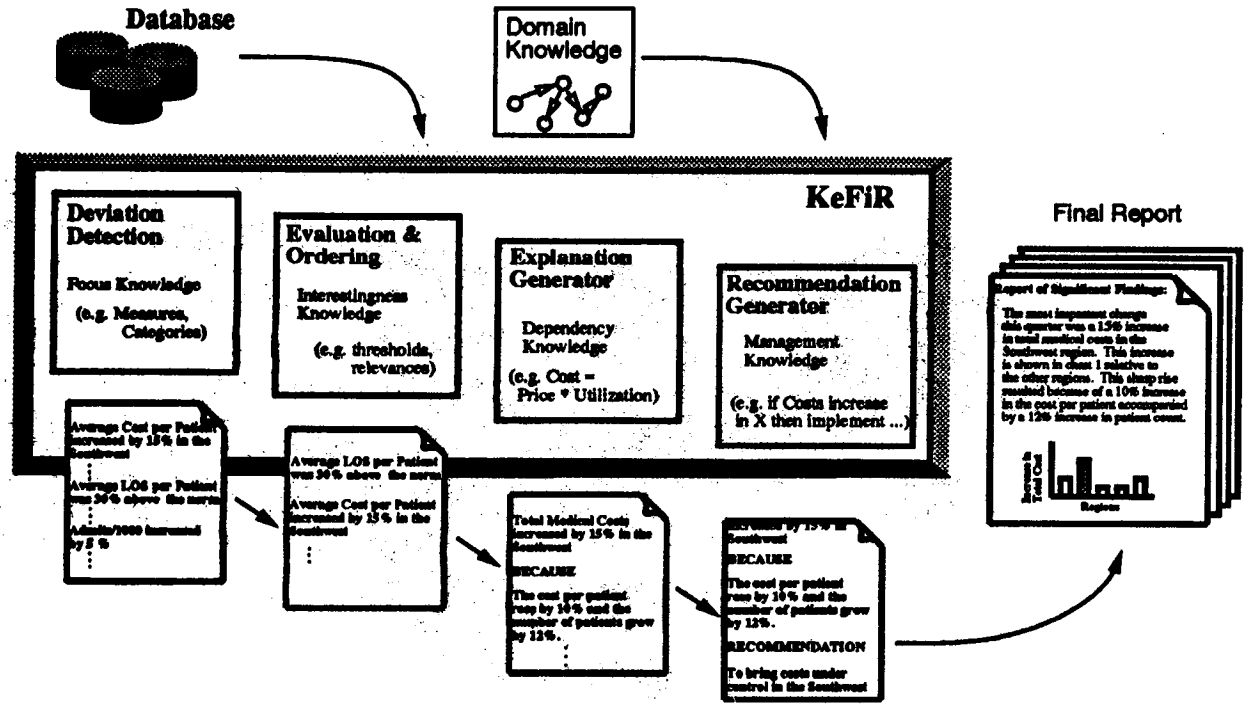

Figure 1: *Overall design and process flow within* KEFIR.

ordering and selecting the most important ones for reporting. Furthermore, a deviation alone is often unsatisfying without some explanation for why it occurred. In the area of health care at least, an analysis is incomplete without some recommendations for how to remedy the problems represented by the discovered deviations.

KEFIR performs all four of these tasks: deviation detection, evaluation, explanation, and recommendation. The overall design and process flow of the system is depicted in Figure 1. Raw data from a database and domain knowledge extracted from the experts are the two sources of input. The system calculates relevant deviations, evaluates and orders them according to their degree of salience, generates explanations for the most important deviations, and recommends courses of actions where appropriate. The final result is a written report with text, tables, and charts.

4.1 Deviation Detection

The deviations that KEFIR explores are completely specified by predefined measures and by predefined categories used to create sub-populations. We refer to a population (or sub-population) as a *sector*, with the "top sector" representing the entire population covered by the data. KEFIR begins its analysis by evaluating the trend and normative deviations of all the measures relevant to the top sector. New sectors are then created for each of the partitions defined by all relevant categories, and deviations are calculated for each measure in each of these new sectors. This drill down into smaller and smaller populations continues recursively until a pre-specified depth is reached or the size of a population becomes inconsequential. The result of this detection process is several hundred to several thousand deviations.

Deviations are encoded in KEFIR within structures we call *findings*. Each finding stores information for a single measure within a single sector. Both the trend and normative deviations are stored within the finding structure. Additional information is also maintained regarding the the impact of the finding, its relation to other findings (for use in explanation), references to relevant measure and sector information, and miscellaneous book-keeping data.

4.2 Ordering Deviations

After the deviations are calculated, they are ordered in preparation for selecting the *key findings* to include in the final report. This ranking requires a metric for calculating the relative salience or importance of a deviation. The metric used by Health-KEFIR incorporates two principal factors: the impact of a deviation and the "probability of success" associated with the finding's recommendation.

The impact of a deviation is its estimated contribution to the total payments made in the top sector. We often refer to this as the "bottom-line impact" because it is an estimate in dollars of how much the deviation is potentially costing the company. This is particularly evident with normative deviations in which case the impact represents the savings that could have been achieved had the value been equal to the expected norm. These potential savings, which are of great interest to health-care managers, appear in portions of the finished report (see appendix).

A recommendation's probability of success, as specified by the health-care expert, is an estimate of how likely the recommendation's action is to bring the measure back to the norm. This probability is multiplied by the impact to derive a prediction for how much money can be saved if the action is followed. It is this "potential savings" that defines the relative measure of salience used in ranking the list of findings (see [Piatetsky-Shapiro and Matheus, 1994] for more details on this process). The top N findings are then selected as key findings for inclusion in the final report. The value N is set to a minimum of five but can increase depending upon the data. The structure of the final report is such that we always want to report on at least five sectors from inpatient and outpatient care, and so N is increased until this is achieved.

4.3 Explanation

KEFIR generates explanations for all its key findings. An explanation for a given finding can come from the decomposition of a formula that defines the finding's measure, or from the breakdown of the measure into its values from the sub-sectors derived from the finding's sector. The decomposition of a measure by formulas is shown in Figure 2. In this example, the measure `Total_payments` can be decomposed by three different formulas. The factors in these formulas are drivers of the `Total_payments` measure since a change in any one directly affects a change in the value of `Total_payments`. Using this knowledge, we can begin to explain an observed deviation in `Total_payments` by relating it to the factor most responsible.

The breakdown of a sector into sub-sectors is illustrated in Figure 3. The high level `Inpatient` sector can be broken down into sub-sectors by several different categories. The highlighted category in this example, `Admission_type`, breaks the `Inpatient` sector into

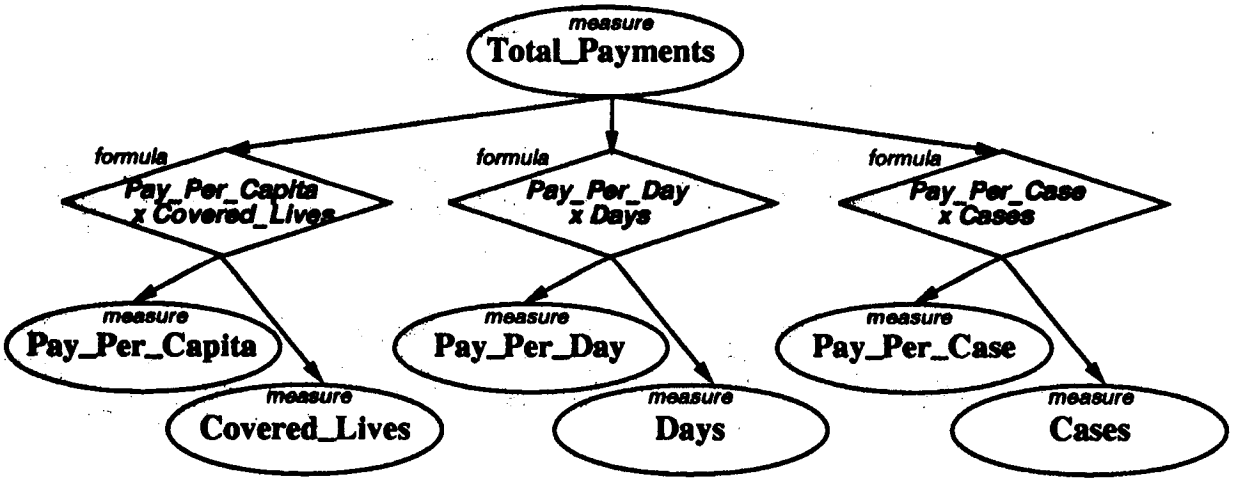

Figure 2: *An example of how measures are related by formulas. The cause of a deviation in* **Total_payments** *can be traced through formulas to deviations in other measures.*

four disjoint sub-sectors. If a deviation is observed in a measure, such as **Total_payments**, we can determine which if any of these sub-sectors is most responsible by comparing their own deviations for that measure. Although this example shows but a single breakdown, in practice there may be many levels, resulting in increasingly smaller and more homogeneous sub-sectors of the population.

KEFIR explains a key finding by first evaluating all other findings affecting it through formulas or breakdowns. It then selects the one finding with the greatest influence and attempts to explain it in the same manner. This recursive process continues until there are no more interesting findings to explain. The final result is a sequence of explanations that chain together a set of interesting findings.

4.4 Recommendation

The main purpose for reporting the key findings is so something can be done to improve the delivery of health care. In many cases, the information provided by a finding is sufficient for the system to automatically suggest an appropriate course of action for handling the problem. **Health-KEFIR** uses a set of rules to identify these situations and to generate recommended actions. The following is the content of a simple recommendation rule:

```
IF measure = In_adms_per_1000 &
   sector = Catastrophic      &
   percent_change > 0.10
RECOMMEND "The increase in catastrophic admission rate suggests a
   review of early preventive methods."
WITH probability of success = 0.4
```

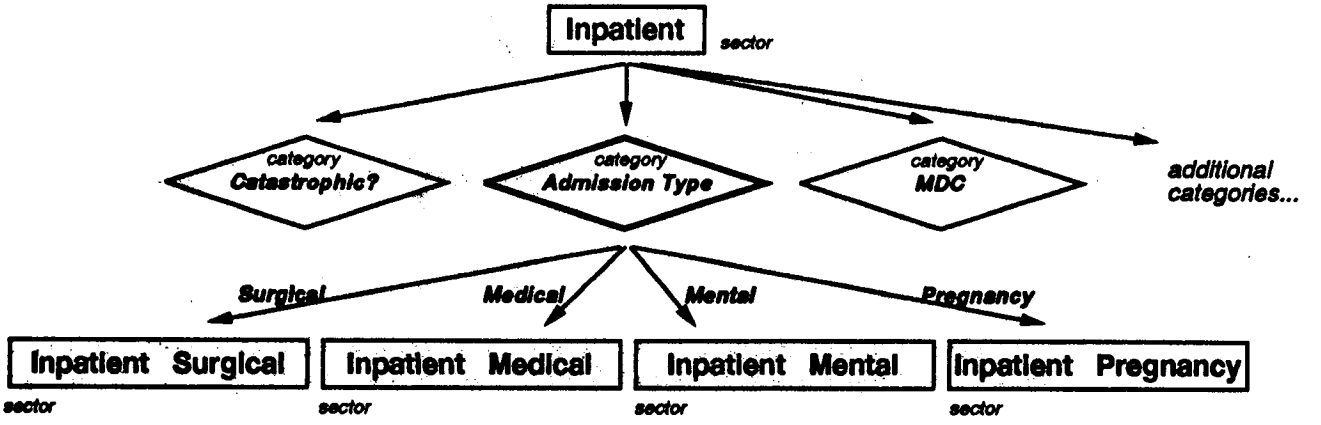

Figure 3: *Examples of how a sector is broken down into sub-sectors using predefined categories. The resulting tree of dependent sectors is used to explain deviations by tracing the value of a measure down into the sub-sectors contributing the most to the observed deviation at the high level.*

4.5 Report Generation

The final output from KEFIR is a written report of the key findings, their explanations, and recommendations. Sentences and paragraphs are generated using simple template matching, with randomized variations to produce more natural sounding text. Descriptive information relevant to the findings also appears in the report in the form of tables, bar charts, and pie charts. To facilitate generation of output appropriate for various word processors and presentation tools, an intermediate file is produced of the report's unformatted content. Formatted sample output for portions of a report generated by Health-KEFIR appears in the appendix.

5 The Application

KEFIR was written entirely in tcl [Ousterhout, 1990] and C/C++, in order to make it portable to virtually all computer platforms. The system's access to data is implemented through an SQL interface which ensures conformance to a wide range of database servers. We are currently running the system on a SPARCstation 10 with an Informix DBMS; we plan a wider deployment on 486 PC's accessing several gigabytes from an Oracle server. The design and development of KEFIR required approximately one man-year. Another four man-months went into the knowledge engineering required to construct the knowledge base for Health-KEFIR. The bulk of this knowledge is represented in a collection of instances of base sectors, categories, and measures. Figure 4 shows parts of the structure definitions for typical instances of a sector, category, and measure.

Health-KEFIR performs its analysis on a central workstation, but it makes its results available remotely by creating a collection of HTML (hypertext markup language) and GIF (graphic interchange file) files and serving these over the network using NCSA's httpd (hypertext transfer protocol) server. The information manager for GTE's Managed Health Care

```
Sector: In_medical_admission
  name:         {medical admissions}
  categories:   {MDC}
  sqltemplate: {where "ADMTYP = 2"}

Category: Admission_type
  name:      {admission type}
  split_by: ADMTYP
  values:    {value: 1 sector: In_surgical_admission}
             {value: 2 sector: In_medical_admission}
             {value: 3 sector: Pregnancy_admission}
             {value: 4 sector: In_behavioral_admission}
             {value: 5   name: in_other sqltemplate: {where "ADMTYP = 5"}

Measure: In_ALOS
  name:      {average length of stay}
  type:      use
  units:     days
  weights:   3
  format:    %.2f
  compute_as: {sql_code {select "AVG(DAYS)" from INPCASE}}
```

Figure 4: *Samples of three of the structures used in* Health-KEFIR. *The slots shown in these examples are only those that have predefined values, i.e. they represent elements of domain knowledge. Additional slots exist for each structure, the values of which are filled in at run time.*

Program accesses the reports using NCSA's Mosaic, a WWW (world-wide web) client for displaying HTML documents. From there the report can be printed for wider distribution or copied into local files for editing into specialized reports.

The initial deployment of **Health-KEFIR** took place in early 1994. Deployment to GTE's regional mangers across the country is scheduled for second half of 1994.

6 Benefits

Much of the time and money corporations spend for health-care analysis is focused on processing the data, quality control, and relatively simple descriptive reporting. Most of this analysis amounts to no more than counting and sorting functions, with precious little attention given to sophisticated analysis or to good consulting advice on recommendations for interventions. **Health-KEFIR** can supplant the "manual" traditional reports that are produced by benefits consulting and health-care information firms. This automation will reduce data-analysis costs by huge amounts, and should provide a solid foundation for health-care consultants to analyze and interpret the resulting information.

In addition to price advantages and better deployment of scarce resources, there are also speed advantages. The typical report may take several weeks to complete manually – with **Health-KEFIR** these reports can be done in a few hours. This permits a more timely response to discovered problems. It also increases the manager's willingness to request reports that would have otherwise been too expensive to justify or to long in preparation to be of use. This is particularly important in a large, complex company like GTE where there are so many organizational subgroups (combinations of business units, geography, union vs non-union employees, etc.) which cannot receive dedicated attention because of resource scarcity. **Health-KEFIR** expands the scope of analyses by making it feasible to produce more of these reports for a wider range of study groups.

Health-KEFIR also promises to improve the accuracy and completeness of reports. The system is not prone to human error, and it can perform a more thorough search than a human expert would have the time or patience for. Human experts typically execute a top-level drill-down, following paths that continue to show promise, while pruning most others for lack of time. **Health-KEFIR** can perform an exhaustive search of all paths through the data to ensure that it does not miss any significant findings. We have observe this kind of thing in practice. In one of its early reports, **Health-KEFIR** identified an important key finding concerning an excessively high re-admission rate for surgical patients – this finding was missing from a report on the same study group submitted by a respectable consulting firm.

In summary, **Health-KEFIR** promises to lower costs, reduce production time, improve accuracy/completeness, and increase coverage of possible study groups. Its early use here at GTE supports these claims. More reports are being generated on a wider range of study groups than was previously possible. The cost per report is infinitesimal compared to the $10,000 plus price tag of a consultant's report, and the turn-around time has been on the order of a day or less. The savings to GTE on report generation costs alone are anticipated to be on the order of several hundred thousand dollars in 1994. Other benefits, such as savings from earlier intervention or from discoveries the consultant firms might have missed, are

more difficult to measure but are perhaps even more important to the long term reduction of cost and improvement of quality of care.

The market for a Health-KEFIR system is enormous. In a company like GTE, the analytic cost savings are estimated at 25% of total health-care information costs. Perhaps of more importance is the market of small and medium-size employers which has never been penetrated because of the fixed initial costs.

7 Limitations and Extensions

The performance of Health-KEFIR is only as good as its domain knowledge. We are adding to its rule base to broaden and improve its recommendation capabilities. New categories and measures are being added to reflect changes in the way health-care managers want to aggregate and decompose information. The weights on measures and sectors are fine tuned as required to meet the expectations of the managers – the automation of this process through the use of a learning algorithm is an anticipated future extension.

Currently the system only handles simple trend and normative analysis. We would like to extend its capabilities to include trend analysis over multiple periods, and add model-based comparisons. With longer range trend analysis it will be possible to modify the salience of a measure to reflect the importance of a constant trend towards or away from the norm. For example, if a measure compares favorably with the current norm but has been steadily increasing while the norm has remained reality constant, this may be a cause for attention which the current system would miss. While norms are useful references for average performance, it is often desirable to set other targets for comparison. For example, rather than comparisons to the average it has been argued that comparisons should be made to a "best practice" model, i.e. comparison to a target representing an achievable level of above average performance. In some situations unusual circumstances may make even the average unachievable, in which case we might wish to set sights on some target below the norm. To accommodate these forms of analysis and tracking, Health-KEFIR will need to be able to use models to represent deviation reference values.

As the complexity of Health-KEFIR advances to meet the complexity of the domain, it becomes more difficult to accurately measure a finding's relative salience. In the current implementation the salience function is a rather simple combination of trend and normative impacts multiplied by the measure and sector weights. We have considered alternative functions that can account for trend directions or involve more complicated interaction between impacts, but making these fit the experts' knowledge has been challenging. Adding more forms of deviations to a finding will certainty add to the salience function's complexity.

Another information product that KeFiR may provide, which is new to the health-care information market, is an "early warning report". This report would monitor many key indicators over many subgroups and would flag variations from expected levels, for further study by benefits managers. This capability, along with other more interactive extensions, are being explored for the next version of Health-KEFIR.

8 Conclusions

KEFIR is a system for rapidly developing discovery applications in domains where trend and/or normative analysis is appropriate. The successful implementation of Health-KEFIR demonstrates the merits of the system and the potential power of the deviation methodology. This technology has matured to a point where wider application to other domains is now feasible and desirable. Within GTE, systems for analyzing data in the areas of marketing, customer support, and telephone operations are being considered.

Acknowledgments: We are very grateful to Shri Goyal and Bill Griffin for their encouragement of our work on discovery in databases.

References

[Anand and Kahn, 1992] T. Anand and G. Kahn. SPOTLIGHT: A data explanation system. In *Proc. Eighth IEEE Conf. Appl. AI*, 1992.

[Matheus *et al.*, 1993] Christopher J. Matheus, Philip K. Chan, and Gregory Piatetsky-Shapiro. Systems for knowledge discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 1993.

[McNeill, 1993] Dwight McNeill. A comprehensive set of performance measures to evaluate managed health care organizations: GTE's perspective. In *National Quality Management Conference*, December 1993.

[Ousterhout, 1990] John K. Ousterhout. TCL: An embeddable command language. In *Proceedings of the 1990 Winter USENIX Conference*, pages 133–146, 1990.

[Piatetsky-Shapiro and Frawley, 1991] G. Piatetsky-Shapiro and W. J. Frawley, editors. *Knowledge Discovery in Databases.* AAAI/MIT Press, Cambridge, MA, 1991.

[Piatetsky-Shapiro and Matheus, 1994] Gregory Piatetsky-Shapiro and Christopher J. Matheus. The interestingness of deviations. In *Proceedings of AAAI-94 KDD workshop*, July 1994.

[Piatetsky-Shapiro, 1993] Gregory Piatetsky-Shapiro, editor. *Workshop Notes from the Eleventh National Conference on Artificial Intelligence: Knowledge Discovery in Databases*, Washinton, DC, July 1993.

[Schmitz *et al.*, 1990] J. Schmitz, G. Armstrong, and J. D. C. Little. CoverStory – automated news finding in marketing. In *DSS Transactions*, pages 46–54, Providence, RI., 1990. Institute of Management Sciences.

APPENDIX: Samples from a Health-KEFIR report

KEY FINDINGS

The following items highlight the key findings in this report:.

- Payments per case in inpatient care increased by 31%, from $9,162 to $12,047, a value $2,996 in excess of the expected norm of $9,050. The hospital discount needs to be examined to see whether there is a potential price problem. A potential savings of $1.6 million would have been possible if this value had been equal to the norm.

- Payments per case in surgical admissions increased from $14,818 to $23,187 (56%), an amount 51% above the expected value of $15,345. A study is suggested of discretionary and high-cost surgery. If payments per case had been equal to the normative value, the savings would have been $1.2 million.

...

SURGICAL ADMISSIONS

Total payments in surgical admissions increased by 46%, from $2.2 million to $3.2 million, due most significantly to an increase in price. One major reason for this was that payments per case in surgical admissions increased from $14,818 to $23,187 (56%), which was $7,842 above the expected norm of $15,345. A study is suggested of discretionary and high-cost surgery.

Surgical admissions				
	1990	1991	% Change	Norm
Total payments	$2.2 million	$3.2 million	46.0%	–
Payments per case	$14,818	$23,187	56.5%	$15,345
Payments per day	$2,336	$3,285	40.6%	$2,536

...

Inpatient Care MDC Breakdown

A breakdown of total payments by major diagnostic categories shows the following principal drivers: