2. **SPECIFIC AIMS:**

Patients’ records and data collected routinely in hospital offer a rich source of information about patient’s condition, patient’s treatment and patient’s responses. The wealth of data collected and available nowadays can be utilized in a variety of ways: data visualization, decision-making, predictions, etc. Our research focuses on explorations of novel possibilities of clinical data repositories: the analysis and detection of clinical outliers (anomalies) and their utilization in patient management and knowledge discovery. To demonstrate the benefits of anomaly detection we propose to study their application to monitoring of patient-management decisions and evaluation of their quality and appropriateness.

Despite numerous improvements in the health care practice over the past decade, the occurrence of medical errors remains a persistent and serious problem [Kohn 1999, Starfield 2000]. The urgency and the scope of this problem prompt the development of solutions aimed to aid clinicians in eliminating such mistakes. Current computer tools for monitoring patients are primarily knowledge-based; the ability to monitor depends on the knowledge represented in the computer and extracted a priori from clinical experts. However, such an approach comes with two disadvantages: the systems are very time consuming to build and they are less robust to missing or incomplete knowledge, that is, their applicability does not extend or generalize beyond what is explicitly represented in their knowledge base.

To alleviate some of the difficulties of knowledge-based approaches we propose to study and evaluate a qualitatively new monitoring and error detection framework that relies on stored clinical information of past patient cases and on statistical methods for identification of clinical outliers (anomalies). We believe that the detection of anomalies corresponding to unusual patient management decisions or outcomes would help to identify potential medical errors. Such a framework may complement the deployment of a knowledge-based system and improve its robustness or it may work as a standalone system in a sub-area in which knowledge-based systems are not available.

The key idea behind our approach is that past patient data offer an excellent source of information of standard patient management practices. The deviations from these standard patient management patterns then indicate unusual choices which may signal a potential medical error. As an example, consider the heparin induced thrombocytopenia (HIT) [Warkentin 2003a] in post-surgical cardiac patients [Walls 1992, Warkentin 2003b]. The condition, if left untreated, could develop into thrombosis (about 50% of untreated patients) or death. The HIT is detected by changes in platelet count and treated by immediate discontinuation of heparin therapy, the choice of non-heparin anticoagulant and other measures. Observing a patient on heparin who has a platelet drop corresponding to the condition, but who continues to receive heparin, would then represent a deviation from prevailing patient management practices and a clinical anomaly. Alerting a physician to such an anomaly would lead to the identification of a potential patient-management error and its elimination. We anticipate that many such patterns can be induced from data.

**Specific Aim 1:** Evidence-based anomaly detection framework.

Anomaly detection methods can be very useful in identifying unusual events or patterns [Hawkins 1980, Barnett 1994, Eskin 2002, Cooper 2004]. We propose a new probabilistic anomaly-based detection framework, called **CADS (Conditional Anomaly Detection System)** that can automatically identify patient-management decisions in the patient record that are highly unusual with respect to past patients with the same or similar condition. The framework builds upon recent advances in machine learning and probabilistic modeling, in particular Bayesian networks, and combines them with techniques applied in statistical anomaly detection and hypothesis testing. The method is designed to complement existing knowledge-based detection methods that are clinically precise, but costly to build. The approach draws its power from a large dataset of past clinical cases; and it requires minimal manual knowledge-acquisition cost.

**Specific Aim 2:** Evaluation of the framework on decisions problems for post-surgical cardiac patients.

To investigate the new anomaly detection framework, we propose to evaluate it on discharge, medication and lab-order decisions made for post-surgical cardiac patients. This cohort is at risk of developing a variety of life-threatening conditions, including the heparin induced thrombocytopenia (HIT), that should be closely monitored and promptly managed. We will use data for past patient cases to evaluate the ability of the anomaly detection system to automatically extract and detect the deviations of patient management decisions (1) for patients at risk of HIT and (2) for the general population of cardiac post-surgical patients. The performance of the system on these tasks will be evaluated by a panel of clinical experts.

We are confident that the state-of-the-art anomaly detection methods we propose in this work are not restricted to intelligent monitoring and alarming systems. We envision these methods, when developed, to be applied in knowledge discovery, where surprising outcomes or relations are uncovered from patient records. For example, we can identify an unusual (positive or negative) response of a patient to a certain type of a treatment. The detection of such a relation, after its subsequent validation, might lead to identification of interesting patient cases or subgroups, which in turn could lead to improved patient-management procedures and knowledge-based rules.