Artificial Intelligence in Medicine 50 (2010) 75–82

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

# Detecting 'wrong blood in tube' errors: Evaluation of a Bayesian network approach

Jason N. Doctor [a,*], Greg Strylewicz [b,c]

[a] Department of Clinical Pharmacy & Pharmaceutical Economics & Policy, School of Pharmacy, University of Southern California, 1540 East Alcazar Street, CHP-140, Lost Angeles, CA 90089-9004, United States
[b] University of Washington, Seattle, WA 98195, United States
[c] Medicine/Northwest Lipids Research Laboratories, 401 Queen Anne Avenue North, Seattle, WA 98109-4517, United States

ARTICLE INFO

*Article history:*
Received 26 August 2008
Received in revised form 20 April 2010
Accepted 26 May 2010

*Keywords:*
Patient safety
Bayesian networks
Autoverification
Laboratory medicine
Medical errors

ABSTRACT

*Objective:* In an effort to address the problem of laboratory errors, we develop and evaluate a method to detect mismatched specimens from nationally collected blood laboratory data in two experiments.
*Methods:* In Experiments 1 and 2 using blood labs from National Health and Nutrition Examination Survey (NHANES) and values derived from the Diabetes Prevention Program (DPP) respectively, a proportion of glucose and HbA1c specimens were randomly mismatched. A Bayesian network that encoded probabilistic relationships among analytes was used to predict mismatches. In Experiment 1 the performance of the network was compared against existing error detection software. In Experiment 2 the network was compared against 11 human experts recruited from the American Academy of Clinical Chemists. Results were compared via area under the receiver-operator characteristic curves (AUCs) and with agreement statistics.
*Results:* In Experiment 1 the network was most predictive of mismatches that produced clinically significant discrepancies between true and mismatched scores ((AUC of 0.87 ($\pm$0.04) for HbA1c and 0.83 ($\pm$0.02) for glucose), performed well in identifying errors among those self-reporting diabetes ($N = 329$) (AUC = 0.79 ($\pm$0.02)) and performed significantly better than the established approach it was tested against (in all cases $p < .0.05$). In Experiment 2 it performed better (and in no case worse) than 7 of the 11 human experts. Average percent agreement was 0.79 and Kappa ($\kappa$) was 0.59, between experts and the Bayesian network.
*Conclusions:* Bayesian network can accurately identify mismatched specimens. The algorithm is best at identifying mismatches that result in a clinically significant magnitude of error.


## 1. Introduction

Medical errors are a significant problem in the United States. They kill more Americans each year than motor vehicle accidents, breast cancer, and AIDS combined [1]. In laboratory medicine, of particular concern are patient identification errors. Proper patient identification is essential to reducing errors and improving patient safety. The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) recognizes this and has included "Improve the accuracy of patient identification" as one of its "National Patient Safety Goals" [2]. Patient identification and other laboratory errors have received increased attention in the research literature both inside [3] and outside [4,5] the United States. In this paper we propose a method that can be used to screen for an error that is particularly difficult to identify in the laboratory, the mislabeled specimen or "wrong blood in tube" error. This type of error refers to a specimen of blood collected on Patient A, but for which the accompanying requisition and label is for Patient B [3].

A wrong blood in tube error is more pernicious than many other blood laboratory errors. If a patient's results are like most others in the lab, then a mismatched sample will often yield a result that is similar to that of the patient's true result. Further, for any set of values for which a proportion of specimens are switched statistical characteristics (e.g., mean and standard deviation) will be the same as if they were not switched. In sum, to identify such mismatched specimens, more sophisticated methods are needed than simple comparisons of values to a norm. In this paper, we develop, train and test a network for detecting wrong blood in tube errors when glucose and HbA1c analytes are analyzed in separate vials. We report on two experiments. Experiment 1 evaluates the network against an established method for automatic detection of errors, LabRespond [5], using the National Health and Nutrition Examination Survey (NHANES) data set. Experiment 2 compares the performance of a Bayesian network to expert lab reviewers when values are derived from a pre-diabetic population.

* Corresponding author. Tel.: +1 323 442 3435; fax: +1 323 442 1462.
*E-mail addresses:* jdoctor@pharmacy.usc.edu, jdoctor@usc.edu (J.N. Doctor).

0933-3657/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2010.05.008

### 1.1. Bayesian networks

We approach the problem of detecting wrong blood in tube errors by implementing a Bayesian network [6]. A Bayesian network is a graphical representation of a joint probability distribution over a set of random variables. A Bayesian network $B = \langle G, P \rangle$ consists of a representing graph, $G$, and an associated joint probability distribution, $P$. The graph, $G$, in the network is described by a finite set of nodes $V$ and a binary relation, $R$ on $V$. A binary relation on a set of nodes $V$ is a subset of ordered pairs $(v_i, v_j)$ in $V \times V$. The relation $R$ characterizes edges in the graph, where $R = \{(v_i, v_j) \in V \times V: v_i \text{ is a } \textit{parent of } v_j\}$. Let $v_i R v_j$ denote $v_i$ is a *parent of* $v_j$. The relation $R$ is *irreflexive* (for every $v_i \in V$, not $v_i R v_i$) and *acyclic* (for any finite sequence of distinct elements $v_1, v_2, \ldots, v_k \in V$ such that $k > 1$ and $v_j R v_{j+1}$ for all $j \in \{1, 2, \ldots, k-1\}$, not $v_k R v_1$). An irreflexive graph is called a *directed graph*, its edges are directed edges, and thus, graphs in Bayesian networks are referred to as *directed and acyclic graphs* (DAGs). The DAG, $G$, in the Bayesian network, $B = \langle G, P \rangle$, represents the probability distribution, $P$, where nodes in $V$ characterize random variables and directed edges describe stochastic dependence. If $v_i$ is a variable in the graph then the graph specifies conditional probability distributions $P(v_i|\pi(v_i))$, where $\pi(v_i)$ are parents of $v_i$. While each variable $v_i$ is dependent on its parents, it is also conditionally independent of any of its non-descendants given its parents. Hence, given a directed acyclic graph $G$ with a set of nodes $V = \{v_1, \ldots, v_n\}$ the joint probability distribution of the network may be factored as follows:

$$P(v_1, \ldots, v_n) = \prod P(v_i|\pi(v_i)) \quad (1)$$

### 1.2. Bayesian networks and blood laboratory errors

Bayesian networks provide a graphical means for representing uncertain relationships between and among variables and allow us to model what might influence belief in why a particular analytic value is observed. In the case of Bayesian networks for detecting blood laboratory errors, we must consider both continuous variables (e.g., analyte values) and discrete variables (e.g., wrong blood in tube: *true* or *false*). To guarantee exact computation, we impose on the DAG the condition that discrete variables are not allowed to have continuous parents [7].

From a network, we may infer a probability that there is wrong blood in the tube given empirical information about observed analyte values and the structure of the network. To understand our approach, consider the following model:

The graph in Fig. 1 encodes knowledge about what influences our belief in analyte values, mismatch and diabetes status. For example, we know of three factors which would influence belief in an (unobserved) HbA1c score:

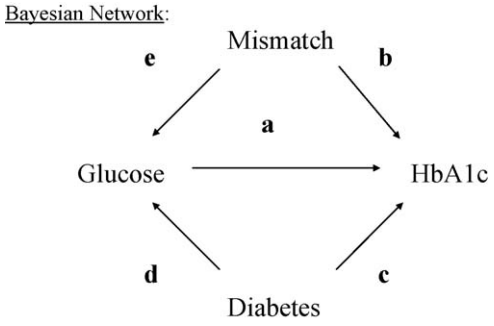


**Fig. 1.** Bayesian network.

(1) Observed glucose score (directed edge "a"), because HbA1c is formed in the patient via a non-enzymatic pathway by hemoglobin's normal exposure to glucose,
(2) knowledge of a mismatch (directed edge "b"), because this, too, may cause one to observe a particular HbA1c score, and
(3) diagnosis of diabetes (directed edge "c"), because HbA1c scores are in general higher for these patients.

Also, disease status (diabetes = ('yes' or 'no')) affects our belief in an (unobserved) glucose score (directed edge "d"). In practice, given both a purported glucose and HbA1c score on a patient, we cannot uniquely identify whether a mismatch was due to a glucose vial switch or an HbA1c vial switch, only that a mismatch in at least one of the two vials may have occurred. Therefore, we draw an arrow from "mismatch" to "glucose" as well (directed edge "e").

Notice in Fig. 1, that absence of arrows communicates important information. For example, our model presented above does not have an arrow from "diabetes" to "mismatch". This is because these events are probabilistically independent. A lab technician handling vials is not more prone to mismatch a diabetic patient's vial as s/he is to mishandle a non-diabetic patient's vial and there is no clear way to justify such an arrow. Therefore, the model we use imposes that one's disease status does not influence belief in a mismatch, but does influence belief in observed fasting glucose and HbA1c score. We note also that to implement the model does not require that diabetes status be known. Belief in diabetes status, however, will be influenced by glucose and HbA1c score. This is an important point, because in a clinical laboratory patient diagnosis is often unknown. The graph in Fig. 1 then constrains the relationships among conditional probabilities among the variables and this network is the basis for our analysis of NHANES data because it incorporates many of the basic facts about glucose and HbA1c analytic results and variables that may influence glucose and HbA1c score.

## 2. Experiments

### 2.1. Experiment 1

#### 2.1.1. Overview

This experiment compares the performance of the network against a validated benchmark method of error detection, LabRespond.

#### 2.1.2. Methods

*2.1.2.1. Data source.* The current study utilized data from the National Health and Nutrition Examination Survey (NHANES). The National Health and Nutrition Examination Survey is an on-going survey and examination of the civilian, non-institutionalized U.S. population. The study is characterized by a complex stratified multistage probability survey design [8]. Mobile examination centers are used for a majority of the health examinations and specimen collections for subsequent analysis at a clinical laboratory. Data from the 2003 to 2004 survey years were utilized in this analysis with glucose from the biochemistry profile, included 6492 results, and glycohemoglobin from the glycohemoglobin profile, included 6601 results. We excluded patients with missing glucose or glycohemoglobin results, leaving a total 6486 patients. Each patient's self-reported diabetic status was incorporated from the medical conditions questionnaire.

*2.1.2.2. Specimen collection.* In order to measure glycohemoglobin, a whole blood sample was collected from the patient by the mobile examination center staff, which then shipped the sample to the University of Missouri-Columbia for analysis using a Primus