

Predictive Analytics and Forecasting in Health Care:

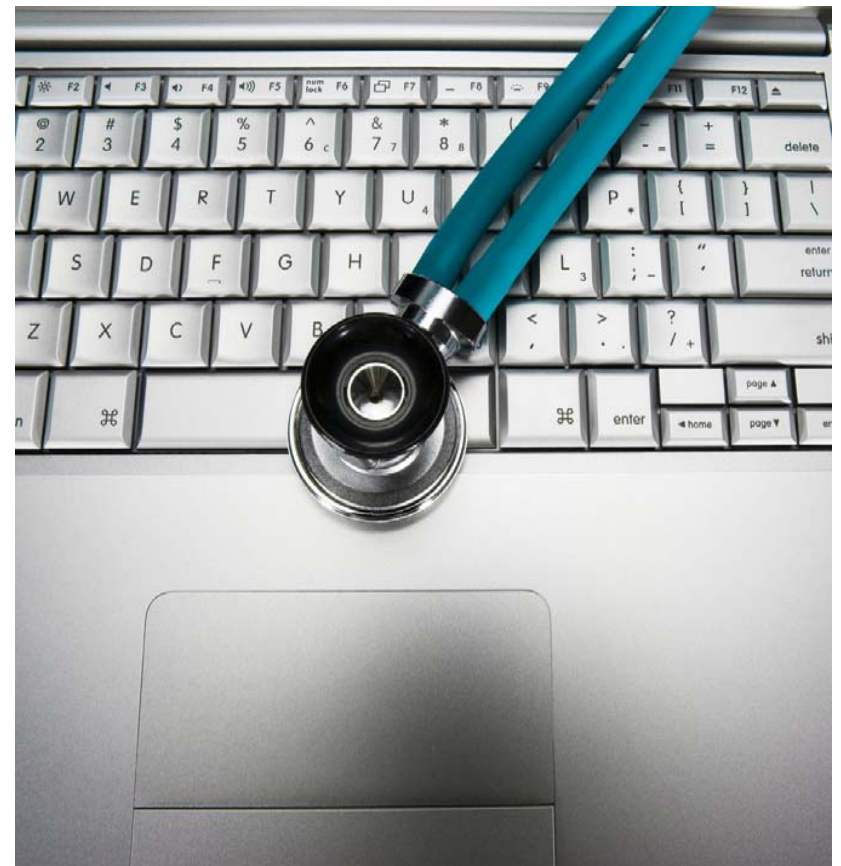
Integrating Analytics with Electronic Health Records

André W. Kushniruk, Ph.D.
Associate Professor and Director,
School of Health Information Science,
University of Victoria

**THE
POWER
TO KNOW®**

Predictive Analytics

- Is the branch of data mining concerned with forecasting of future probabilities
- Why data mining?
 - Goes beyond reporting what has happened
 - Discover why this has happened
 - Power to predict what will happen
 - Use of large amounts of diverse data to uncover new relationships



What Can We Predict?

- Identify patients and populations at risk
- Adherence
- Healthcare utilization
- Identify adverse drug events
- Improve selection of candidates for patient-centered interventions
- Health outcomes
- Identify costly procedures, waste and delays
- Safety issues and risks

Opportunities for Predictive Analytics in Health Care

- Huge amounts of health data now being collected electronically
 - Electronic patient records
 - Major national initiatives and databases
 - Linking patient data across hospitals, regions etc.
- Ability to collect range of data on patient access to their own data and other health information
 - Personal health records
- Large number of health databases
 - Hospitals, health regions, governments
- Need to be able to work across these varied data to:
 - Identify patient and population health needs
 - Optimize and streamline the healthcare system
 - Evaluate information system effectiveness

Electronic Health Records

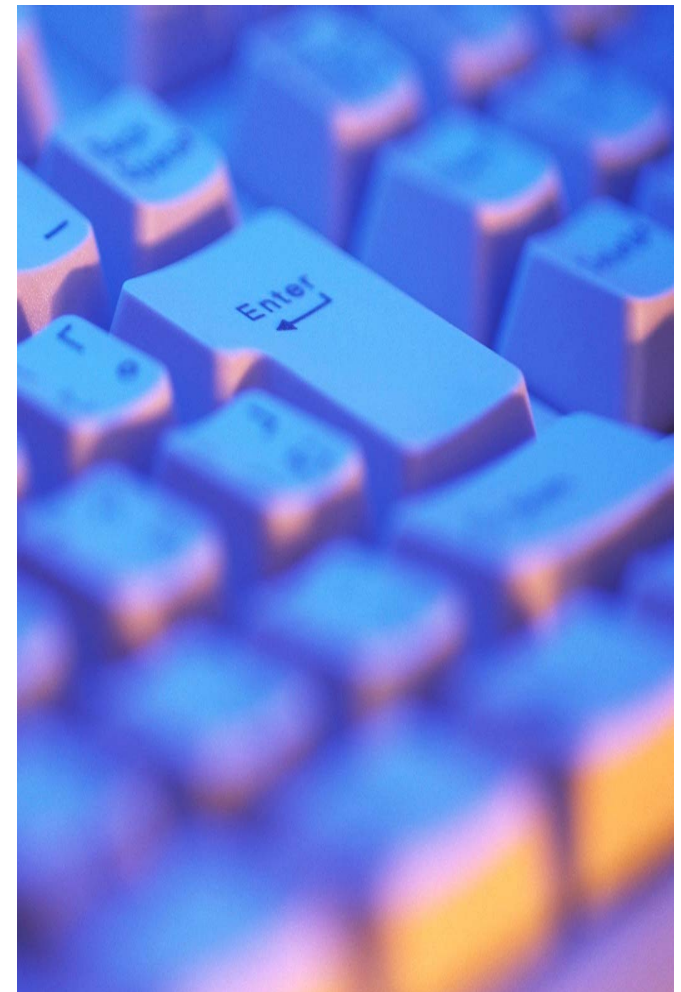
- A repository of electronically maintained information about an individual's health status and health care, stored such that it can serve the multiple legitimate users of the record

EHR: the longitudinal record of an individual that contains data from multiple EMRs and EPRs (shared and/or interoperable across settings)

EPR: the electronic record in a hospital or facility

EMR: the electronic medical record in a physician's office

PHR: The electronic record containing personal health information entered and maintained by the patient or layperson



Adoption of the Electronic Medical Record (EMR)

- Survey – 6 in 10 BC family doctors feel overwhelmed by patient information
- 58% of GPs and 48% of specialists suggested patient health records were only mostly accurate
- 90% of GPs and 82% of specialists said they would make better-informed decisions if the information they had was more relevant to their decision making

- Programs in BC include PITO
- EHR is cornerstone of Canada Health Infoway
- Adoption rates still relatively low as compared to high adopting countries

Adoption of EHR in Other Countries

- Adoption rates highest in the following countries
 - Denmark
 - Sweden
 - Taiwan
 - U.K.

- These countries have well developed national programs that have allowed for interchange of patient data across regions/countries

- National networks and databases of patient record information have been implemented in several countries
 - Web-based physician and patient access
 - Patient Portal - Denmark

- These countries are beginning to apply value-added analytic functions integrated with the EHR
 - predictive analytics
 - data mining

Issues for Data Mining in Healthcare (when integrated with the EHR)

- **Modern medicine generates huge amounts of data**
 - EHR will provide opportunity for richer more relevant and real-time data analysis
- **However, as amount of data increases, there is a widening gap between data collection and data comprehension**
- **Pressures to**
 - Find better methods for analysis
 - Automating them to facilitate the creation of knowledge to support clinical decision making
 - This is where data mining and knowledge discovery come in

Example: Analysis of Usage of Web-Based EHRs/Patient Portals

- Towards “consumer-driven” health care
- Varied users who interact from various locations
 - Less able to conduct controlled evaluative studies
- Current state-of-the-art
 - track user actions (e.g. clicks) - tells what they do, but not why
 - on-line questionnaires/feedback forms - often not filled in, limited questions
 - interviews - problem that users often do not know what they do
 - Can we integrate analysis of behavioral data with health data?

PatCIS Auth - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Netsite: <https://flux.cpmc.columbia.edu/patcis/enter.cgi> What's Related

Data Review

Vital Signs Blood Sugars Laboratory Laboratory Detail PFT Reports Microbiology

Logout

Data Entry

Data Review

Advice

Education

Help

Comments

CHEM 7 PROFILE 1999-05-08 22:57 Status: final				
NA	143	136-146	mM/l	
NA	TEST	-	-	
K	3.5	3.6-5.0	mM/l	
CL	100	102-109	mM/l	-
CO2	-	-	-	-
BUN	21	7-20	mg/dl	
GLUCOSE	100	70-105	mg/dl	-
CREATININE	-	-	-	
ABC 1999-05-08 16:29 Status: final				
WBC	4.6	3.54-9.06	10 ⁹ /l	-
RBC	10.33	4.00-5.20	10 ¹² /l	-
HGB	5.2	12.0-15.8	g/dl	-
HCT	35.4	35.4-44.4	%	-
MCV	79.0	79.0-93.3	fl	-
MCH	26.8	26.7-31.9	pg	-
MCHC	25.3	32.3-35.9	g/dl	-
RDW	2.3	<14.5	%	-
PLT	35	165-415	10 ⁹ /l	-
MPV	5.3	9.00-12.95	fl	-
BLOOD GAS, ARTERIAL 1999-05-08 16:29 Status: final				
pH	7.25	7.35-7.45	-	-
PCO2	34	32-43	mm Hg	-
PO2	76	72-104	mm Hg	-

Document: Done

Screen of a patient clinical information system (PatCIS) showing data review function

Questions in the Evaluation of e-Health Information Systems

- What type of information do e-Health consumers want?
- Is the information provided useful, helpful?
- How to collect useful data from large number of subjects remotely?
- How to integrate data from multiple sources?
- How to analyze such data from varied data sources to discover usage patterns?



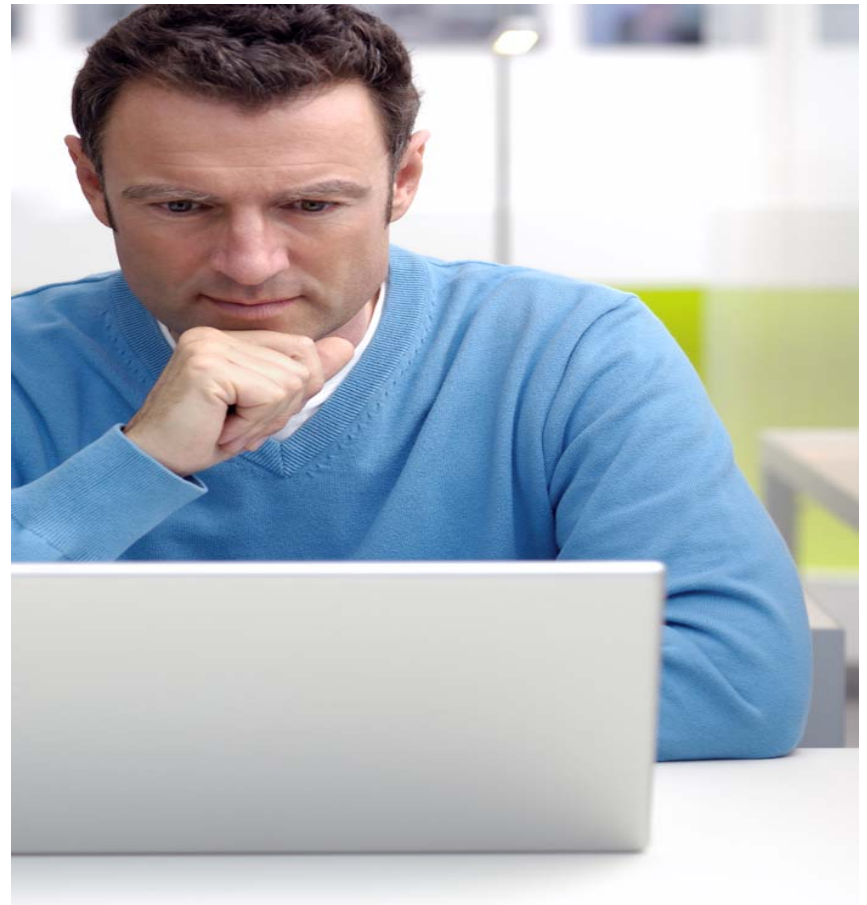
Objectives

- To collect psychologically rich and useful data on a large scale we need:
 - Methods for automatically collecting usability data at point of system use
 - identify patterns of usage of interest to automatically collect data about
 - Analysis tools and discovery tools
 - Automatically identify patterns of usage from merge of data collected

- Goal of Work
 - Integration of multi-method data collection to answer both specific and generic questions regarding usability of Web-based systems
 - Move all forms of usability data to on-line, therefore forming basis for automated analysis

Three Phases

- Phase I - Remote Data Extraction
- Phase II – Preprocessing and Data Merging
- Phase III – Analysis and Pattern Discovery



Phase I – Remote Data Extraction

- Moved laboratory from fixed site to remote data collection

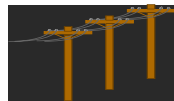
- Varied forms of data collected
 - User interactions with system
 - Health data
 - Questionnaire data
 - Qualitative data



Remote Analysis of System Usage

1. Video Based Usability Testing

- from laboratory to remote



2. Interviews

- from phone to electronic



User

3. E-mail

(to evaluators)

Interact via WWW



4. Tracking User Actions - System Usage Database (log files)

Consumer Information System

5. On-line Questionnaire Data

(triggered forms)

Kushniruk, Patel, Patel, & Cimino, 2001

Types of data collected remotely

1. Collect baseline data on use of a system (tracking using log file) – example of log from part of a single session:

Jones ! Fri Oct 27 11:32:22 2000 ! cpmc.columbia.ed ! PatCis-login

Jones ! Fri Oct 27 11:32:24 2000 ! cpmc.columbia.ed ! PatCis-data-review

Jones ! Fri Oct 27 11:32:28 2000 ! cpmc.columbia.ed ! PatCis-data
review-labs

Jones ! Fri Oct 27 11:32:35 2000 ! cpmc.columbia.ed ! PatCis-help

Jones ! Fri Oct 27 11:33:00 2000 ! cpmc.columbia.ed ! PatCis-logout

Note – log files lead to large databases of usage information. Often difficult to analyze and examined in isolation BUT CANNOT provide direct information about user's subjective experience, intentions etc. for each record

2. Identify “patterns of use” interested in
 - e.g. use of on-line health guidelines on invoking a patient record system by doctor
3. Provide “triggers” for identifying entry into patterns of interest (pattern detectors)
 - Using scripting language
 - Design any computer initiated “dialogues” with end user
4. Continue to log all interactions
5. At trigger points invoke specific on-line questionnaires
 - About usage
 - Explanation of browsing patterns
 - Marketing questions
 - Usability question
6. Patient disease data
7. Fully automated remote recording of user screens and audio

Phase II – Preprocessing and Data Merging

- First stage is data collection (which as described above is automated)
- Second stage is storage of data within knowledge representations that support knowledge discovery



Representational Requirements

- Need for integrated of multi-method data collection and representation on system usability
- Merging of data from multiple sources
 - Log files
 - Questionnaires – baseline, pop-up
 - Usability data
- Collection of large data sets over time
- Need for new methods for analysis of integrated data sets
- Knowledge-based (object-oriented) approach taken
 - To data representation
 - To data analysis

Phase III – Analysis and Pattern Discovery

- Function Usage
 - Most frequently accessed function was “Review of Laboratory Data”
 - Accessed by patients at least once in the majority of the sessions
 - “Review of Reports” was second most frequently accessed function
 - Other functions (advice, education and data entry) were used sparingly



Analysis of User-System Interactions

Statistical Analysis of Usage (number and percentage of accesses)

<u>ADVICE</u>	<u>DATA ENTRY</u>	<u>DATA REVIEW</u>	<u>EDUCATION</u>	<u>TOTAL</u>
.3%	4%	93%	3%	100%

- **Majority of accesses by patients for Data Review**
 - Laboratory details
 - Reports – admit/discharge, cardiology, radiology
- **Discovery of patterns of usage related to both demographic and medical data (association rules)**
 - **Most used and useful for patients with specific illnesses – chronic illness (e.g. diabetes)**
 - Association between nature of illness (chronic) and high level of system usage discovered
 - **Automatically explored relationship of usage to**
 - Education, demographics, illness, medications, perception of interaction with providers (over time)

Pattern Discovery : Range of Analysis Methods

- From simple statistical analysis to
- On a continuum in terms of complexity
- Basic statistical methods and mining for association rules most useful for our purposes



Simple statistical methods

Association rules

Hidden Markov modeling

Discovery of Patterns of Potential Interest

- Association rule mining
 - E.g. Association between disease type (e.g. diabetes) and pages viewed (e.g. laboratory results)
- Clustering
 - E.g. clusters of similar patterns of usage, etc...
- Classification
 - E.g. Classification of users of system by level of expertise etc.
- Sequential patterns
 - E.g. specific web page is visited after accessing a specific patient report
- Usage statistics
 - E.g. functions accessed, related functions etc.
- Application of pre-defined deductive rules across data

Examples of Association Rules Discovered (and Practical Application)

1. Studies of use of clinical guidelines by physicians:

Context of clinical use -> format of guideline accessed/preferred and satisfaction

Patient encounter time -> format of guideline accessed/preferred and satisfaction

- led to development and refinement of different guideline formats for different medical contexts

2. Studies of use of patient clinical information system by patients:

Degree of chronic nature of illness -> level of use of patient clinical information system

- led to refinement of system to target chronic patient information needs

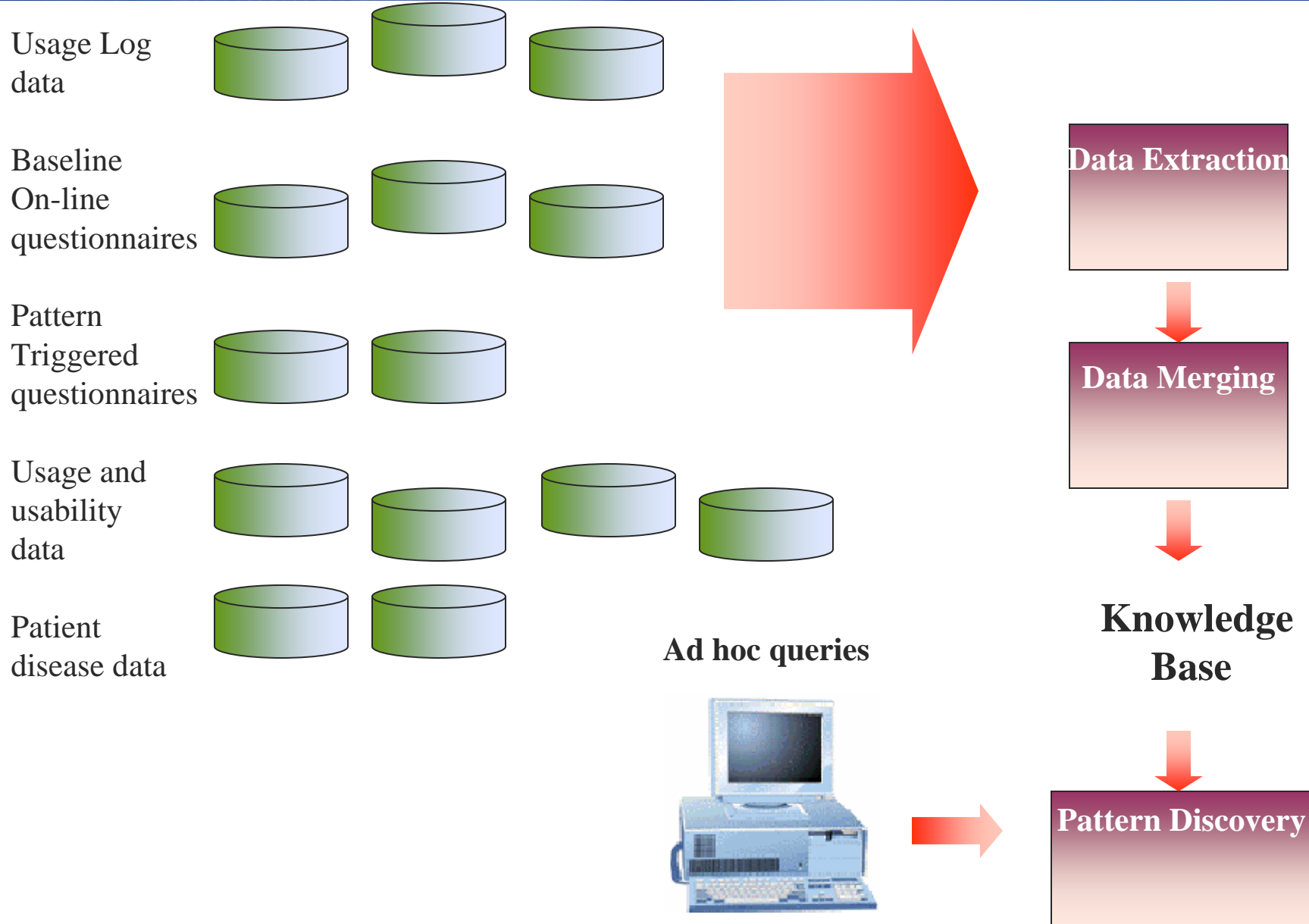
Further Analysis

Ad-hoc querying of merged usability and health data

- E.g. what information do patients need to know about to manage their diabetes and is use of a clinical information system related to improvement?

Application of pattern-detection algorithms - automated pattern discovery tools

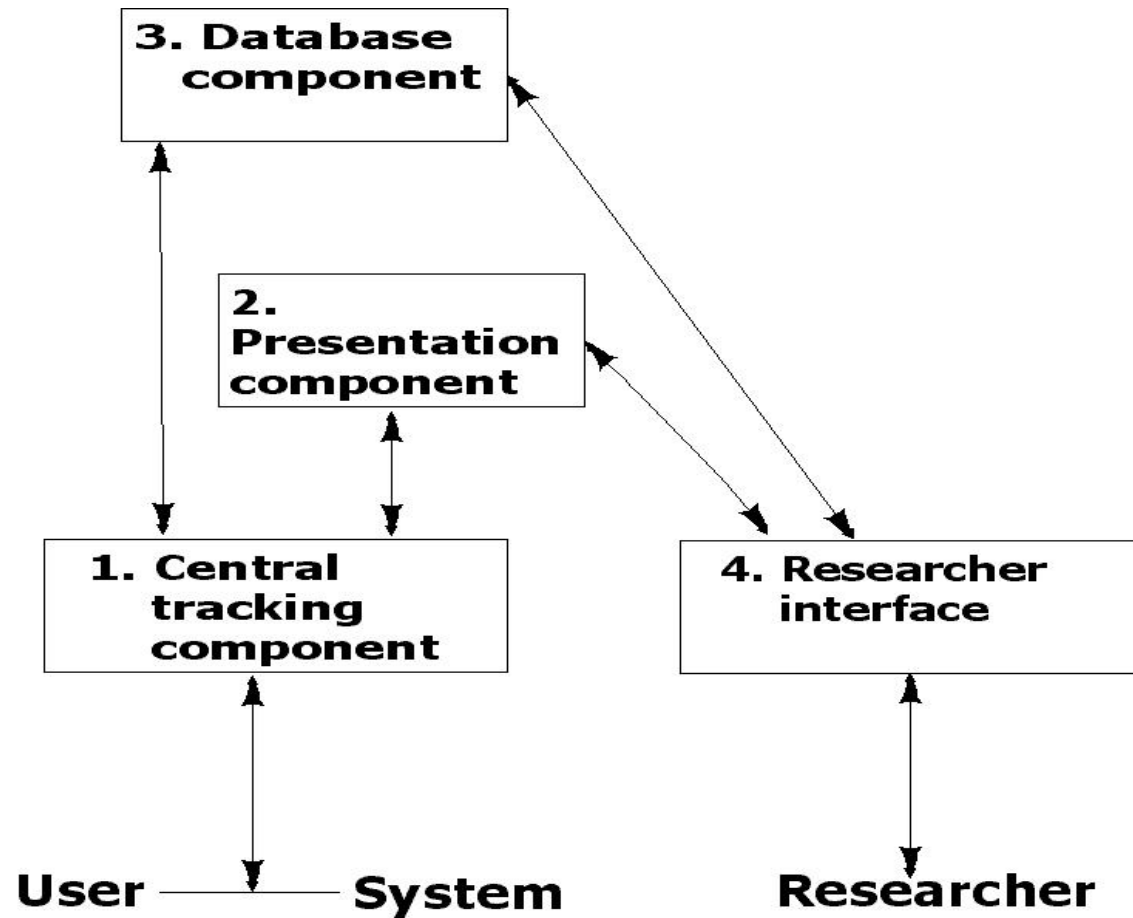
- Need for approaches to data mining large warehouses of data
- Can have important impact on marketing e.g. relation of drug usage to patient exposure and access to health guidelines
- Can link usability data to data in computerized patient records, allowing for integration with health info)



Overall Approach

Towards an Integrated Analytic Toolkit: The Virtual Usability Laboratory (VULab)

- **Objective to develop a research software tool that facilitates automated data collection and analysis**
 - Automated tracking and analysis of multiple forms of data obtained remotely
 - e.g. log files, questionnaires, triggered on-line interviews, remote recordings of computer screens
 - Automated storage of information in relational (and OO) databases to allow for ad-hoc querying
 - SQL queries about how users are interacting with a system
 - Development of a user-friendly interface for researchers and evaluators
 - Allows them to interact with the system in setting up evaluations and selecting analysis methods



Overall VULab Architecture



Figure 5a. Resource page showing links to clinical guidelines available from within a computer-based patient record system.

FEEDBACK FORM - Netscape

File Edit View Go Communicator Help

Location: http://132.206.83.200/cgi-bin/columbia_project/feedback.cgi?0

Instant Message Members WebMail Connections BizJournal SmartUpdate Mktplace RealPlayer

Please Take A Moment To Complete The Following:

Please Indicate the Purpose of Accessing the Guideline:

- To help me care for a specific patient**
- For general knowledge/education**
- Other**

If the answer to the above question was "for a specific patient", please indicate the patient-care setting:

- I am seeing the patient now in an out-patient setting (e.g. clinic, emergency room, private office, etc.)**
- I am seeing the patient now in an in-patient setting (e.g hospital admission)**
- I saw the patient previously**
- Other**

Document: Done

Start Screenc My Computer Vol_second (...) test Eudora Light... FEEDBAC... 4:32 PM

Figure 5b. Form to assess clinician’s reason for accessing a guideline (which appears when the user selects a guideline from the resource page).

CLINICAL GUIDELINE, PART 1, Guidelines... Annals 15 Dec 97 - Netscape

Location: <http://132.206.83.200/cgi-bin/checkpoint.cgi?url=http://www.acponline.org/journals/annals/15dec97/pplyme1.htm>

ACP-ASIM
ONLINE

[Home](#) | [Search](#)
[Site Index](#)

ACP-ASIM Online
[For Internists](#)
[About ACP-ASIM](#)
[Where We Stand](#)
[News](#)
[Interact with ACP-ASIM](#)
[Pharmaceutical In-Site Corner](#)
[Find It](#)

Time and Medicine
4 January 2000
Annals

Annals of Internal Medicine
[Current Issue](#)
[Past Issues](#)
[Library for Internists](#)
[Subscriptions](#)
[Info for Authors](#)
[Reprints & Permissions](#)

Annals of Internal Medicine

POSITION PAPERS

CLINICAL GUIDELINE, PART 1

Guidelines for Laboratory Evaluation in the Diagnosis of Lyme Disease

Annals of Internal Medicine 15 December 1997.
 127:1106-1108.

[Related letter](#)
[Related letter](#)

See also: [Clinical Guideline, Part II](#)

American College of Physicians*

The numbers in square brackets are cross-references to the numbered sections in the accompanying background paper, "[Laboratory Evaluation in the Diagnosis of Lyme Disease](#)," which is part II of this Guideline.

Lyme disease is the most common tick-borne disease in North America. From 1982 to 1994, more than 70 000 cases were reported in North America, most of them in the United States. It is important that clinicians diagnose Lyme disease correctly because efficacious therapy is available and delayed or

Document: Done

Start | Screenc | My Comp... | Vol_sec... | test | Eudora Li... | EXIT FO... | CLINIC... | 4:32 PM

EXIT FORM - Netscape

File Edit View Go Communicator Help

Location: http://132.206.83.200/exit_form.html

Instant Message Members WebMail Connections BizJournal SmartUpdate Mktplace RealPlayer

Please take a minute to complete the following after having accessed the guidelines

For your purposes in using this guideline:

I got all of the information I wanted from the guideline
 I got some of the information I wanted from the guideline

I did not get the information I wanted from the guideline because:

The guideline did not apply to the specific patient I was considering
 I could not find information relevant to what I needed to know
 The guideline may be incorrect
 The guideline is not current
 Other, Please specify

THANK YOU!

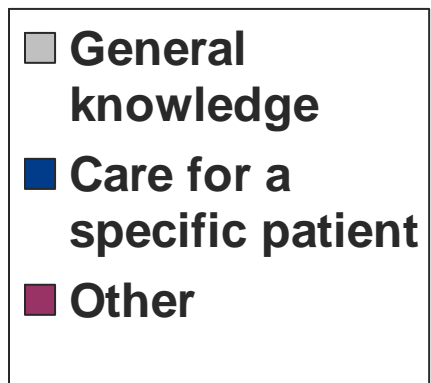
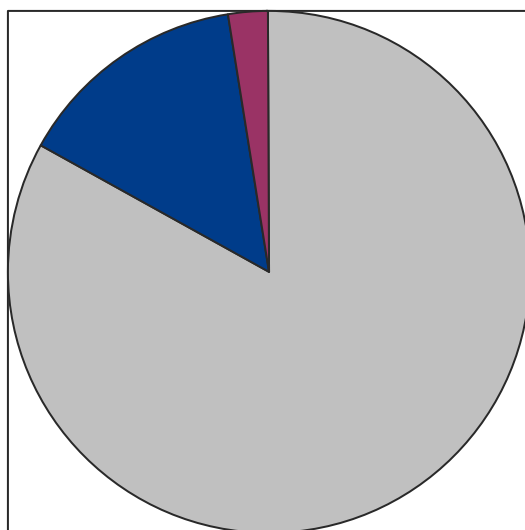
SUBMIT

Document: Done

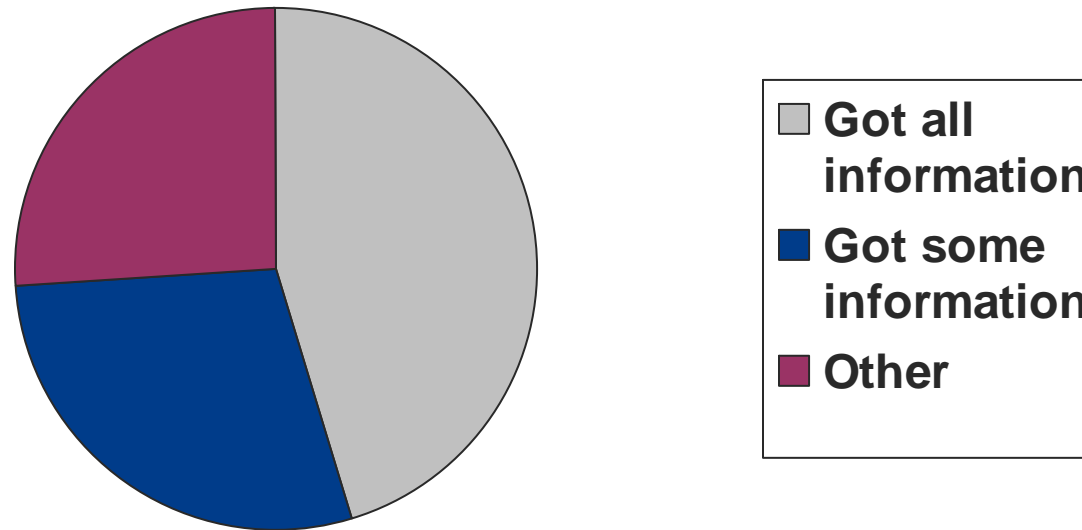
Start Screenc My Comp... Vol_seco... test Eudora Li... EXIT F... 4:33 PM

Figure 5c. Form to assess success in obtaining desired information from the clinical guideline accessed (appears on exit from the guideline).

Context of Use of Clinical Guidelines by Physicians

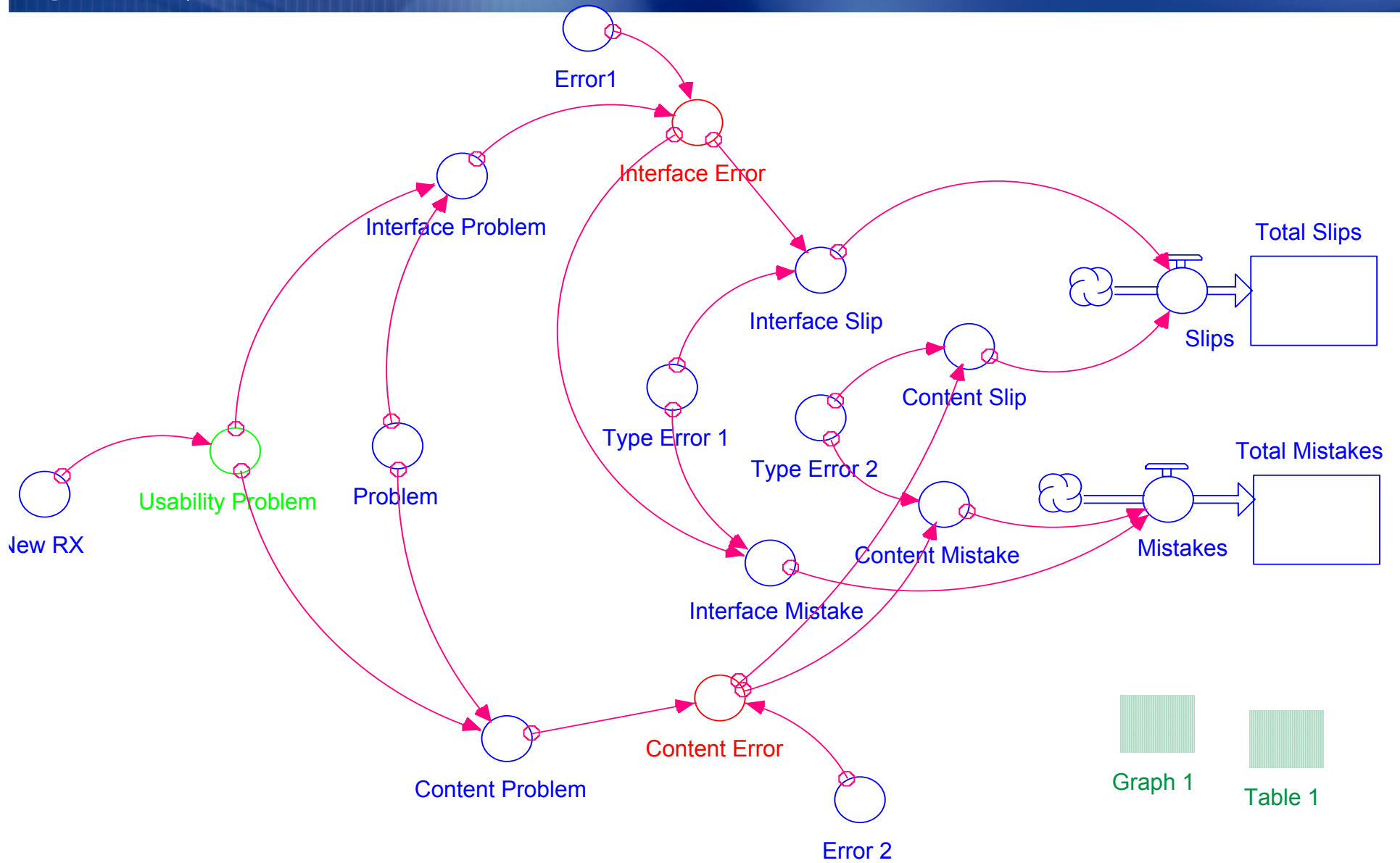


Information Quality for Physicians Accessing Guidelines



Simulation Applications -Predicting Adverse Drug Events

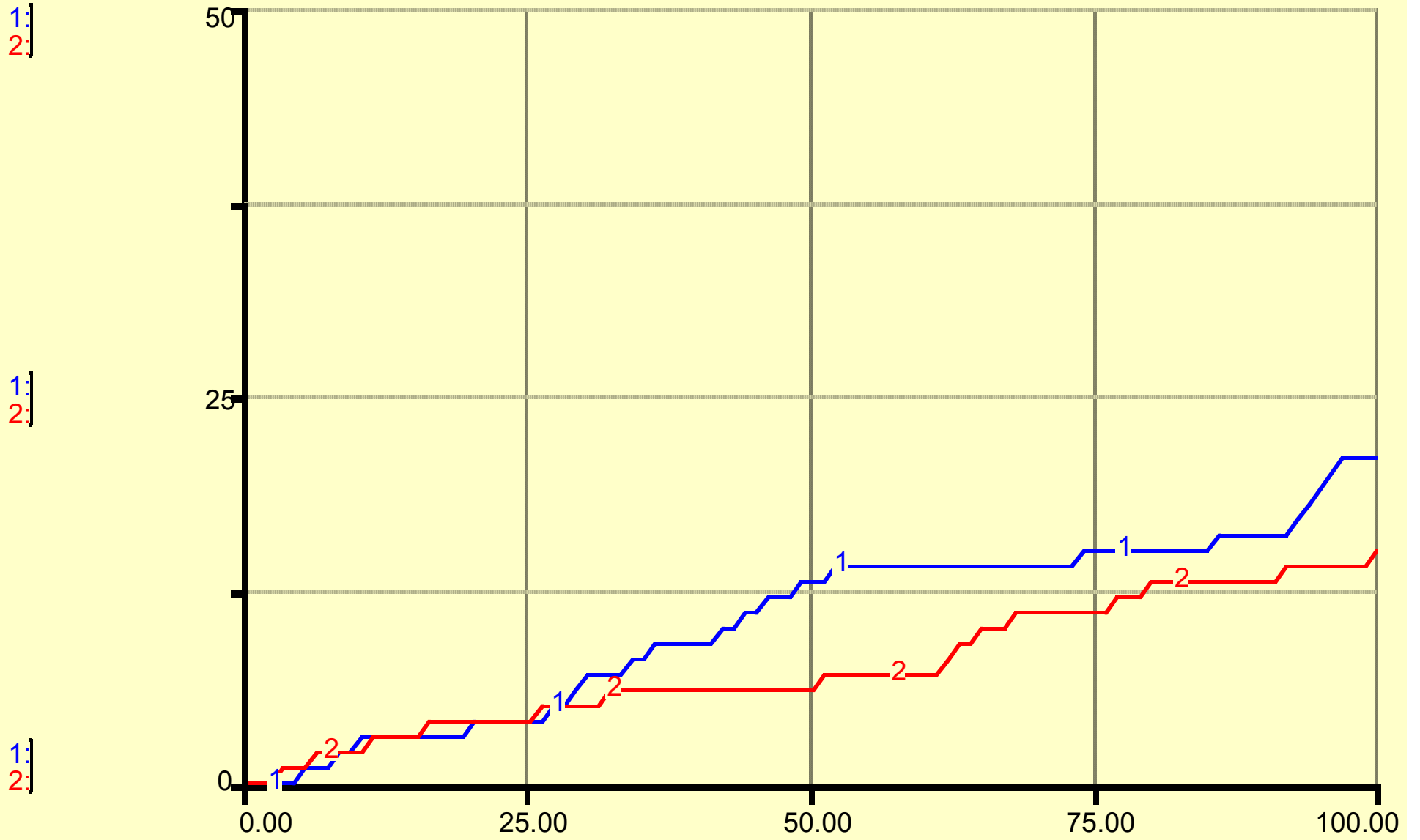
- Extension to modeling and forecasting of long-term and widespread impact (e.g. across health regions) of adverse drug events
- Based on collection of empirical data on error rates (using approaches described above)



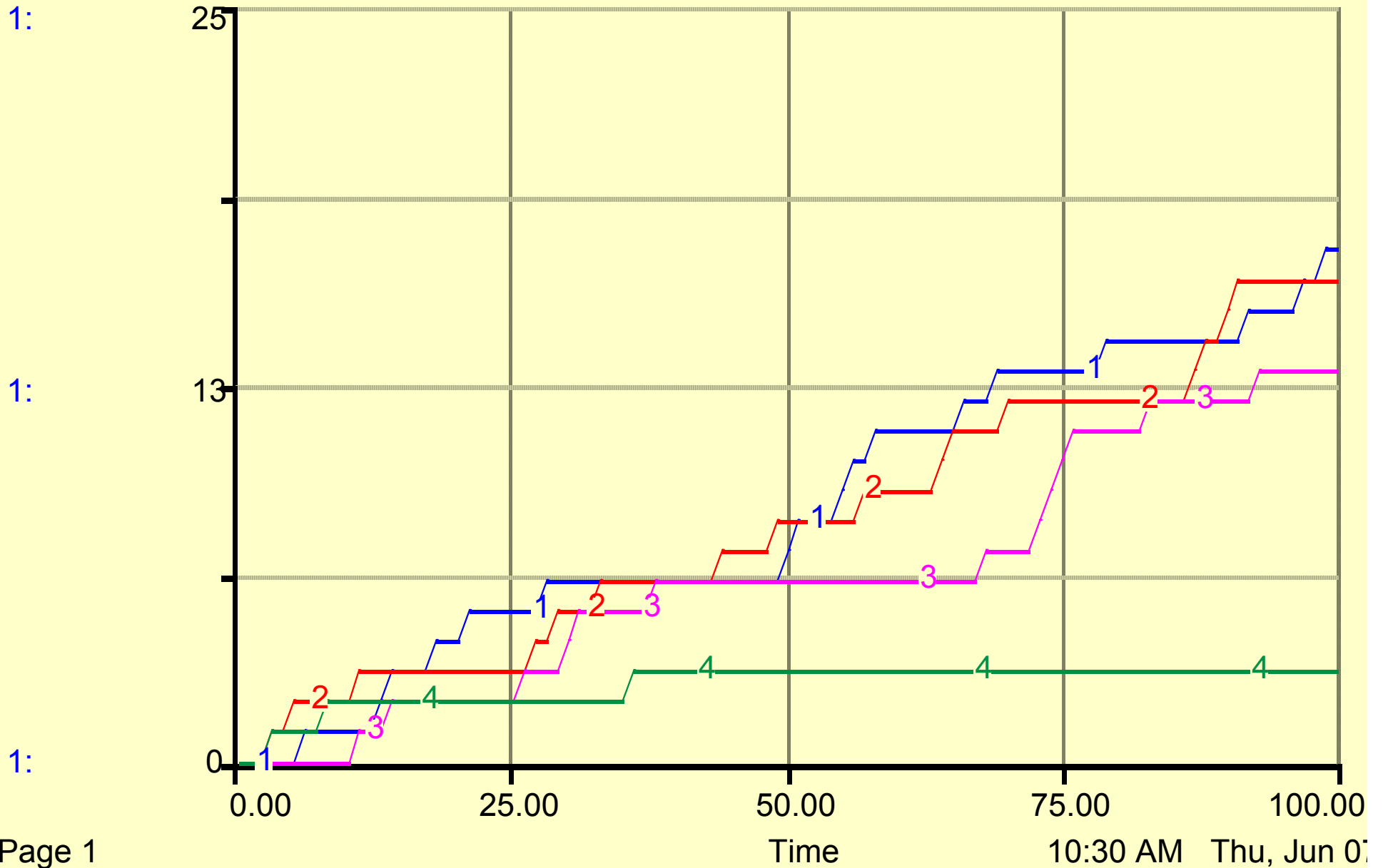
Technology Induced Error and Usability

1: Total Slips

2: Total Mistakes



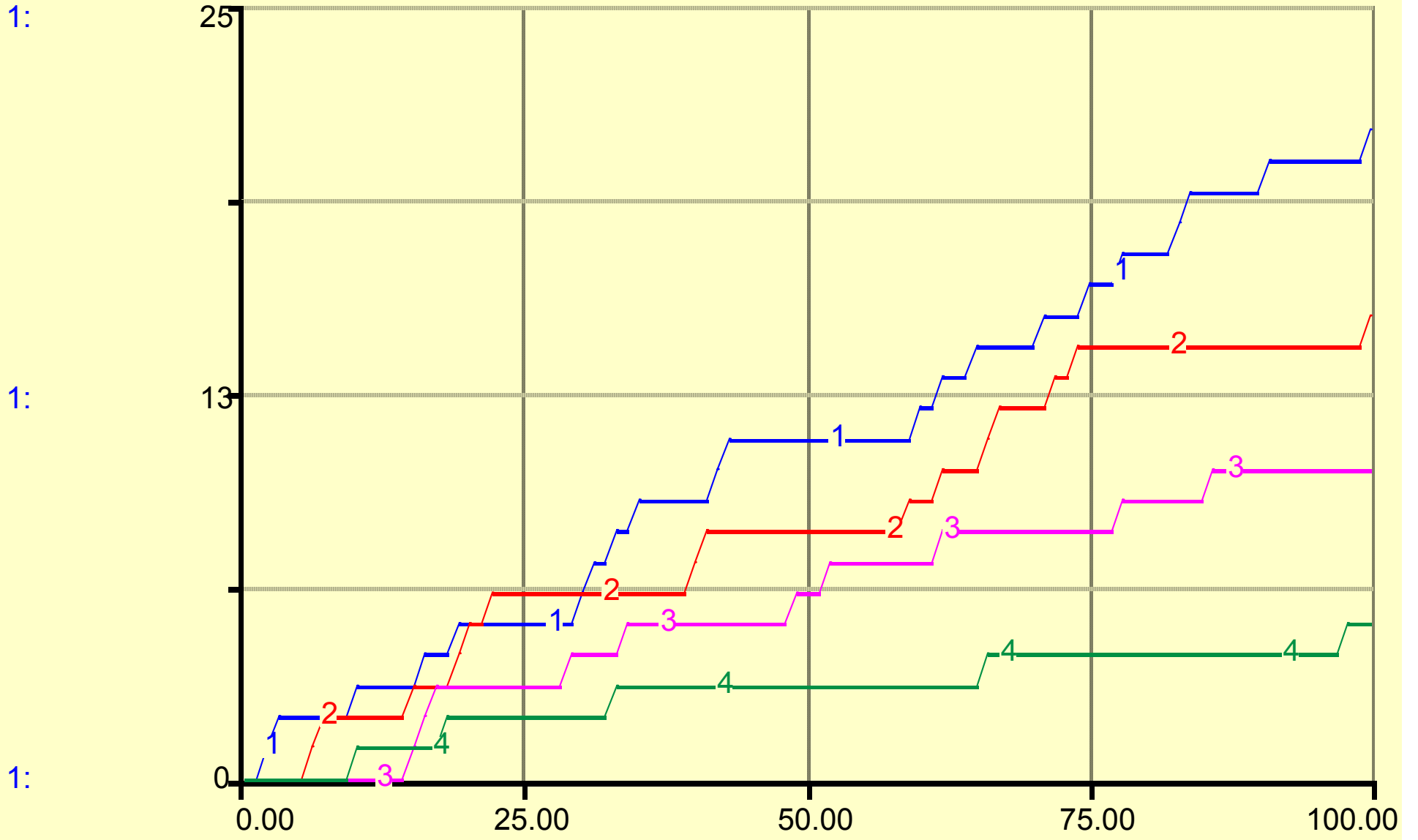
Total Slips: 1 - 2 - 3 - 4 -



Comparative Graph of Total Slips



Total Mistakes 1 2 3 4 -



Comparative Graph of Total Mistakes



The Future of Analytics with the Electronic Health Record

- Applying real-time analytics to electronic health record data
- Provides feedback to physicians decision-makers in real-time to make decisions about patients and patient populations
 - e.g. within an ICU with integrated information system
- Provide health administrators with information in real-time and can predict future health resource needs
 - e.g.
 - staffing of programs,
 - developing programs to address population health needs to prevent disease,
 - developing disease management programs to address chronic diseases in the population

Conclusions

- Next generation of healthcare analytics will be more powerful when integrated with EHR

- Will move from “Evidence-based medicine” to “Analytics-based healthcare”
 - Consumer-driven
 - Dynamic
 - In real time
 - Any time any place
 - Electronic health data more closely tied to quality improvement

- Will allow for future planning involving health care services based on population health needs

Questions?





**THE
POWER
TO KNOW®**