



## Towards a Framework for Data Quality in Healthcare

**Jonathan S Einbinder**

Partners Healthcare System  
jseinbinder@partners.org

**Vipul Kashyap**

Partners Healthcare System  
vkashyap1@partners.org

**Joel L Vengco**

Partners Healthcare System  
jvengco@partners.org

**Executive Summary/Abstract:** Measuring and improving the quality of healthcare delivery are major goals at Partners HealthCare Systems, Inc. Information that can be derived from clinical and administrative information systems (i.e., order entry, electronic medical record, billing information) is core to achieving these goals. We present a Framework for Data Quality in the context of data uses in healthcare. The Framework lends itself to a set of potential solutions for insuring quality data for various healthcare measurements. Case studies were used to illustrate the applicability of the solution approaches in the context of the proposed Framework.

1



## Objectives

- Current State of Healthcare
- Data Quality Issues within Healthcare
- A Framework for Data Quality in the Context of Healthcare
- Conclusions

2



## Current State of Healthcare



## What is the problem?



- Medical errors in 2.9-3.7% of hospitalizations
  - 1/2 preventable
  - 8.8-13.6% lead to death
- 8th leading cause of death
  - More than car accidents, breast cancer, AIDS
- Costs to society \$17-\$29B
  - 1/2 healthcare

Institute of Medicine, 1999

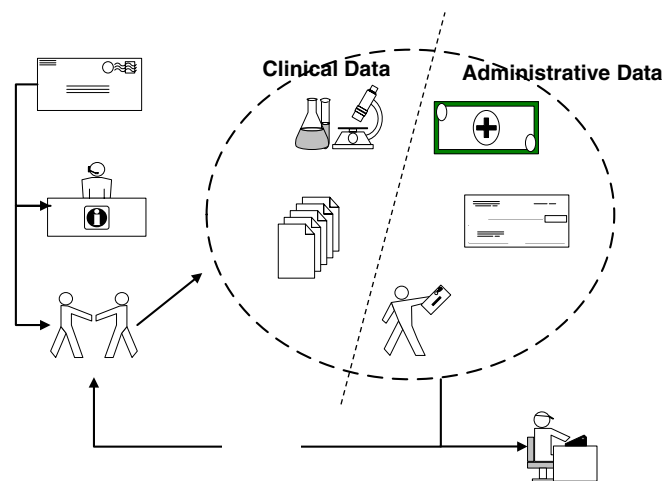


## Needs and Uses for HealthCare Information

	Needs and Uses	Customers
Patient Care	<ul style="list-style-type: none"><li>• Patient/population management</li><li>• Patient record look-up</li><li>• Treatment decision</li></ul>	<ul style="list-style-type: none"><li>• Clinicians at the point of care</li></ul>
Quality Measurement	<ul style="list-style-type: none"><li>• Quality and Safety measurement</li><li>• Quality and Safety improvement</li><li>• Regulatory reporting (JCAHO, CMS)</li></ul>	<ul style="list-style-type: none"><li>• Clinicians</li><li>• Hospital Quality Manager</li></ul>
Research	<ul style="list-style-type: none"><li>• Healthcare/Informatics research</li><li>• Clinical Trials / Controlled</li></ul>	<ul style="list-style-type: none"><li>• Informaticians</li><li>• Clinician researchers</li><li>• Academic researchers</li></ul>
Business Management	<ul style="list-style-type: none"><li>• Utilization reporting</li><li>• Cost reporting and analysis</li><li>• Contract management</li></ul>	<ul style="list-style-type: none"><li>• Executive business managers</li><li>• Department chairs</li></ul>



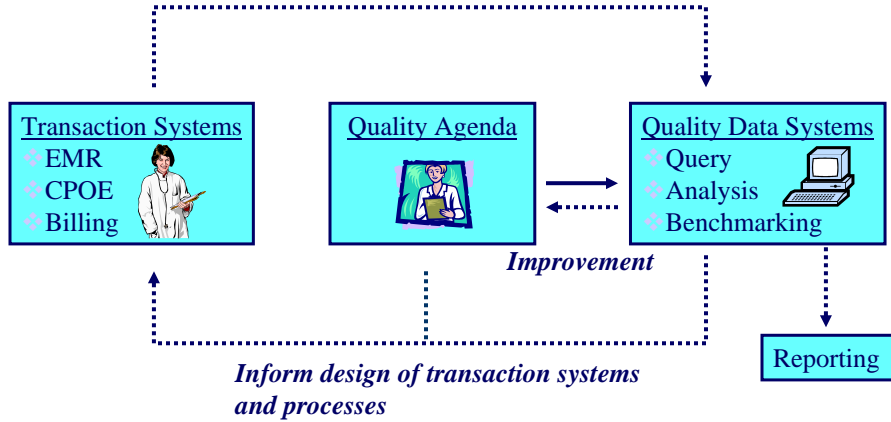
## Sources for HealthCare Data





## Quality Measurement and Clinical Information Systems

*Data from transaction systems  
is extracted for quality measurement*



## Data Quality Issues Within Healthcare

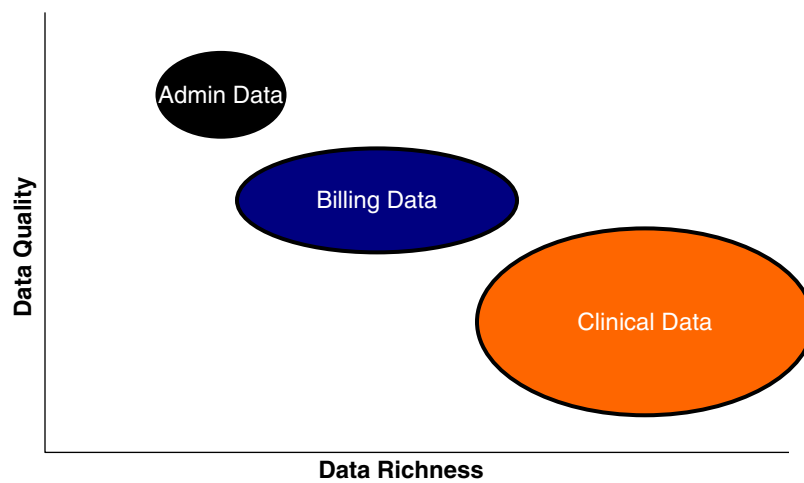


## Characteristics of high quality healthcare data

- Complete
- Accurate
- Source system context
- Codified and standard
- Sufficient level of detail (granularity)
- Data Reusability
  - e.g., data collected at the point of care is reusable for other purposes such as quality analysis



## Characteristics of Healthcare Data





## Data Quality Dimensions in Healthcare

- Data Source Quality:
  - Administrative (high data quality, low data richness)
  - Billing (medium data quality, medium data richness)
  - Clinical (low data quality, high data richness)
- Characteristic:
  - Inherent incompleteness of Medical Knowledge
- Data Semantics:
  - Who are my patients?
  - How severely sick is my patient?
  - What was the reason for the treatment?
  - How do I know who had an eye exam?
  - Compared to whom (what benchmark)?
  - Differing local values/representations (e.g., Med or Lab names, Normal ranges, Abbrevs, etc.)

11



## Data Quality Issues: Inherent Incompleteness of Medical Knowledge

- 250 Diabetes mellitus
  - 250.0 without mention of complication
  - 250.1 with ketoacidosis
  - 250.2 with hyperosmolarity
  - 250.3 **with other coma**
  - 250.4 with renal manifestations
  - ...
  - 250.8 **with other specified manifestations**
  - 250.9 **with unspecified complication**
- 790.6 Hyperglycemia NOS
- 790.2 Nonclinical diabetes
- 648.8 Gestational diabetes

*All medical databases, and medical records, are necessarily incomplete because they reflect the selective collection and recording of data by the healthcare personnel responsible for the patient.*

*-- Shortliffe and Barnett, 2000*

*Current level of coding accuracy and diagnostic precision is insufficient for population-based studies of outcomes of specific conditions or therapies*

12



## Data Source Quality: Administrative and Billing Data

### Advantages

- Readily available
- Easy to capture
- Codified/computer interpretable
- Describes large areas of population
- Standardized across healthcare industry

### Limitations

- Purpose is for reimbursement, no intended for quality assessment!
- Limited Clinical Insight
  - Interpersonal/technical quality of care
  - Error determination, appropriateness
- Limited outcomes
  - e.g. in-hospital death
- Limited Nursing Information
  - e.g. smoking cessation
- Limited reporting of actions, results
  - E.g., Labs, Administration of Meds
- Limited insight on care processes/decisions
  - E.g., Meds administered because of some lab results
- Limited in temporal insights



## Data Source Quality: Clinical Data

### Advantages

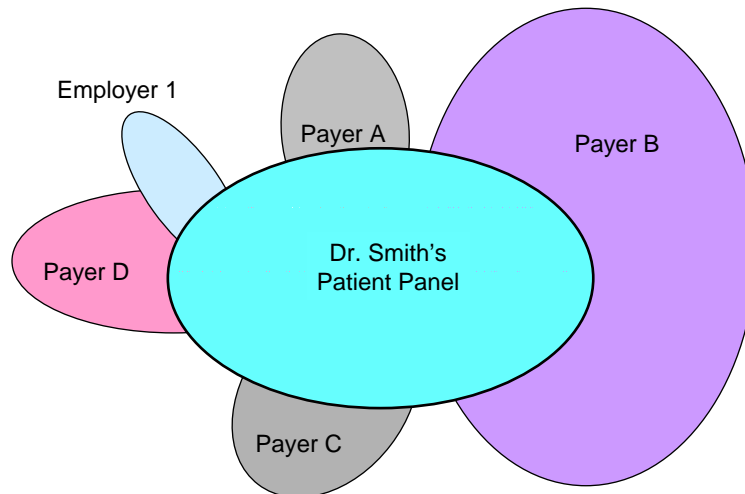
- More accurate representation of patient care (as opposed to administrative data).
- Granular information in clinical data
- Complete account of patient care
  - Meds, Labs, Patient History, Longitudinal Patient Information, Co-morbidities, Family History, Genetic information
- Ability to piece clinical events better based on clinical data

### Limitations

- Unstructured Notes
  - Information locked in Clinical Notes
- Lab values may be represented as free text
- Diagnoses and Findings not codified
- Not all orders have results or are captured
- Administration of Meds, Therapies
  - Unclear capture and representation of what was actually done
- Lack of causality representation in data
  - Did a patient get a test due to an indication?
  - Was a drug administered due to an indication?



## Data Semantics: Cohort Definition



15



## Defining “My Patients” – what is the denominator?

- Who are your patients?
  - Visits (triage, cross coverage)
  - PCP (insurance, reality, ever seen)
  - Intervention and procedures
- Standard definitions are essential
  - Detailed, unambiguous
  - e.g. NCQA's HEDIS measures (<http://www.ncqa.org>)
- Definition used at Brigham and Women's Hospital:
  - A Patient is my patient if:
    - I am listed as the PCP in registration data
    - The patient has visited me more than once in the past 3 years
    - The patient is not known to be dead

16





## A Framework for Data Quality and Solution Approaches



## Data Quality Framework for Healthcare

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements



## Set of Solutions

- Merge data
- Reuse dataset
  - Admin data
  - Non-transformed clinical data
- Merge dataset + Manual prospective entry
- Change process
- Restructure information model
  - Scale normalization



## Data Quality Framework for Healthcare

Re-use non-transformed clinical data

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements

Data Source, Functionality, Data Quality Issue



### Data Quality Framework for Healthcare

Merge Data Set + Manual Prospective Entry

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements

Data Source, Functionality, Data Quality Issue



### Data Quality Framework for Healthcare

Restructure Information Model

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements

Data Source, Functionality, Data Quality Issue



## Case Study 1: Accuracy of Claims Data

- Application Description
  - Quality Measurement: Cohort Definition + Reporting
  - Of my cardiac patients, how many received Cardiac Catherterizations?
  - Billing Data (without change) is appropriate for this; Data Reuse
- Solution Approach
  - Reuse Administrative Data Set
- Solution Characteristics:
  - Administrative databases are a valuable resource collected at great expense
 BUT,
  - Interpret claims-based hospital comparisons with caution
  - Current level of coding accuracy and diagnostic precision is insufficient for population-based studies of outcomes of specific conditions or therapies



## Data Quality Framework for Healthcare

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements



## Case Study 1: Accuracy of claims data (Fisher 1992)

- DRG Validation Study 1985
  - Sample of 239 hospitals
    - Charts reviewed for 7050 discharges
  - Coders assigned ICD-9 diagnosis and procedure codes = GOLD STANDARD
  - Compared with Medicare claims data
- Agreement
  - Similar principal diagnosis 78.2%
  - Similar principal procedure 76.2%
- Sensitivity [principal or secondary dx]
  - 0.58 *peripheral vascular disease*
  - 0.97 *breast cancer*
- Procedures much better sensitivity
  - 0.88 *cardiac catheterization*
  - 0.95 many procedures

25



## Case Study 2: Accuracy of Clinician entered data

- Application Description
  - Using diagnosis codes entered by physicians enhance (or detract) the ability to define and extract:
    - Patient co-morbidities
    - Quality Indicators
  - Quality Measurement: Cohort Definition + Reporting
  - Research: Cohort Definition + Reporting
- Solution Approach
  - Clinical Data and Claims Data; Merged Dataset

26



## Data Quality Framework for Healthcare

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements

Data Source, Functionality, Data Quality Issue

27

Description	Ncase IP	Ncase HSF	% found	Ncase IP only	% not found	Ncase HSF only	% new cases
HIV and AIDS	1192	1016	85.2	176	14.8	121	10.2
Lymphoma	2431	1883	77.5	548	22.5	566	23.3
Psychoses	10921	7295	66.8	3626	33.2	1878	17.2
Congestive heart failure	18512	11185	60.4	7327	39.6	4453	24.1
Solid tumor without metastasis	27986	16802	60.0	11184	40.0	1773	6.3
Liver disease	5094	3004	59.0	2090	41.0	1203	23.6
Weight loss	4047	2116	52.3	1931	47.7	6871	169.8
Metastatic cancer	10596	5355	50.5	5241	49.5	874	8.2
Other neurological					9.9	5169	44.5
Valvular disease					3.0	6717	62.7
Peripheral vascular di					4.3	2988	28.8
Cardiac arrhythmias					9.8	4121	16.2
Renal failure					1.4	751	12.2
Chronic pulmonary di					2.2	2567	10.4
Alcohol abuse					6.4	517	4.7
Drug abuse					7.5	213	5.5
Rheumatoid arthritis/d					3.7	194	5.6
Paralysis					9.9	782	10.0
Coagulopathy					9.9	1127	18.3
Pulmonary circulation					9.9	1520	53.3
Fluid and electrolyte d					0.6	2138	7.5
Diabetes, uncomplica					2.5	3723	15.7
Hypertension					74.0	5094	9.6
Diabetes, complicated	7515	1491	19.8	6024	80.2	450	6.0
Depression	8910	1757	19.7	7153	80.3	2543	28.5
Deficiency anemias	15159	2608	17.2	12551	82.8	2506	16.5
Blood Loss anemia	2087	332	15.9	1755	84.1	315	15.1
Hypothyroidism	10867	1213	11.2	9654	88.8	235	2.2
Obesity	10560	1106	10.5	9454	89.5	183	1.7
Peptic ulcer disease including bleeding	4783	282	5.9	4501	94.1	618	12.9

**Additional cases identified with physician data**

**CHF: 4453 (24%)**

**Depression: 2543 (28%)**

**Weight loss: 6871 (170%)**



True variance?  
This brings you closer to completeness, but not necessarily better quality (e.g., less accurate)

Quality Indicator	# Inpatient visits	Medical Records + Physician Diagnoses		Medical Records Diagnoses only	
		# with complication	Complication Rate	# with complication	Complication Rate
Inpatient Wound Infection	234587	3878	1.7	3305	1.4
Pulmonary Compromise after major surgery	38009	3224	8.5	706	1.9
Acute Myocardial Infarction after major surgery	39381	327	0.8	234	0.6



### Case Study 3: Severity and Risk Adjustment

- Application Description:
  - To make meaningful comparisons among patients, payers, or institutions, must take patient-specific factors (age, sex, comorbid conditions, severity of illness, risk of death etc.) into account
  - Adjustment using only administrative data
    - e.g., **Deyo-Charlson index, APR-DRG**
    - **Physicians don't trust these measures**
  - Adjustment that includes clinical data
    - e.g., **ACC Risk adjustment for Cardiac patient**
    - **Physicians trust these measures**
- Solution Approach
  - Merging Clinical and Administrative Data Sets
  - Restructuring Information Model/Scale Normalization



## Data Quality Framework for Healthcare

	Data Acquisition and Access	Cohort Definition	Reporting
Patient Care	Clinical, Documentation, Unstructured Data	N/A	Clinical, Monitoring and Decision Support, Data Granularity, Unstructured Data
Quality Measurement	Clinical, Billing, Reporting and Documentation, Data Semantics	Clinical, Billing, Reporting and Documentation, Data Semantics, Missing Data Values	Clinical, Billing, Regulatory Compliance, Quality Enhancements, Comparative Analysis, Data Aggregation
Research	Clinical, Admin, Billing, Outcomes identification, Data Semantics, Definition	Clinical, Admin, Billing, "What if" Analysis, Data Semantics, Missing Data Elements	Clinical, Admin, Billing, Comparative Analysis, Data Semantics, Missing Data Elements

Data Source, Functionality, Data Quality Issue

31



## Ongoing Work at Partners HealthCare: Clinical Dashboards for Reporting at Patient Care

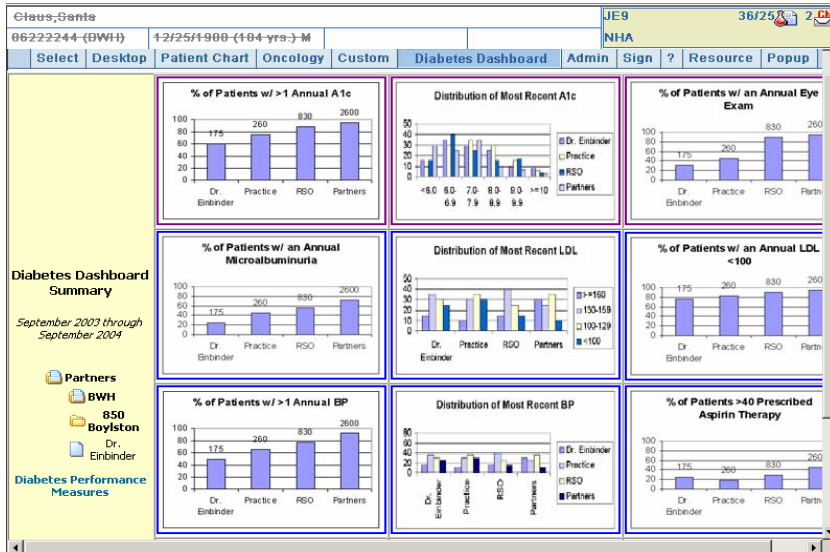
- Implement an online dashboard that describes quality of care with regard to several quality indicators. The intended customer/user of the dashboard is the physician, who will be able to review his/her performance.
  - Ideally, dashboard provides information that is actionable (“population-based clinical decision support”)
  - Touches upon the various data quality issues discussed in the talk.

32

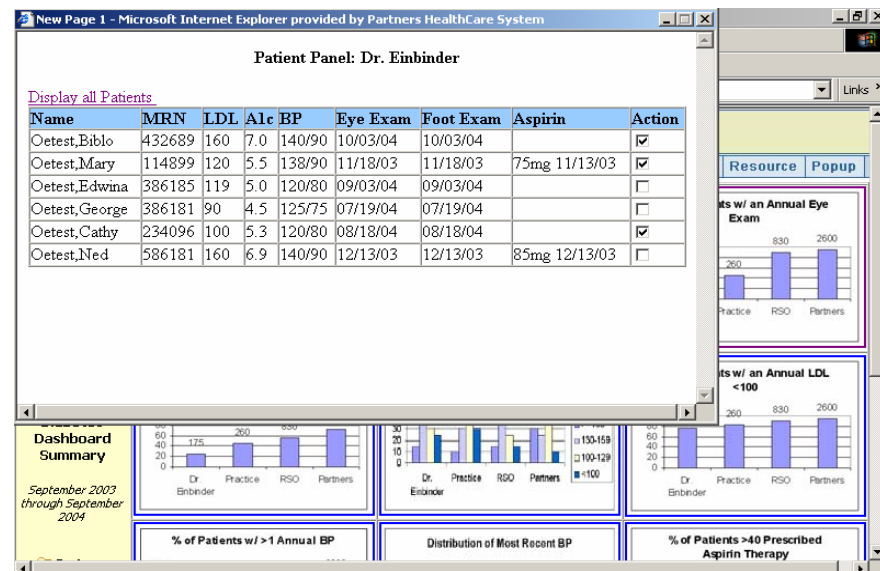




### Clinical Dashboard for Diabetes



### Clinical Dashboard for Diabetes





## Future Work:

- Quality and integration
  - Set and meet specific targets for IOM priority areas
  - Deploy electronic clinical programs system-wide
  - Improve pharmaceutical decision-making
  - Enhance “system-ness” and care coordination

### Opportunity

- Measurement systems
- Registries and cohorts
- System-wide data and infrastructure
- Clinical Decision Support
- Population decision support
- Patient computing

35



## Conclusions

- Leveraging the connections between enterprise information systems and the quality of healthcare delivery and practice
  - Clinical Information System
  - Billing and Administrative Information System
- Using data for various Healthcare initiatives (e.g., measure healthcare quality)
  - Administrative Data: Easy to extract, less accurate
  - Clinical Data: Difficult to extract, more accurate
  - Need to leverage both to estimate healthcare quality
- Proposed a Framework for Data Quality in Healthcare
- Proposed initial solution approaches in the context of the Framework to enhance data quality
- Presented Case Studies that exemplify the use of the Framework

36