



# Big Data and Health Care

Jed Weissberg, MD, FACP  
Senior Vice President, Hospitals, Quality and Care Delivery Excellence  
Kaiser Foundation Health Plan and Hospitals  
Kaiser Permanente

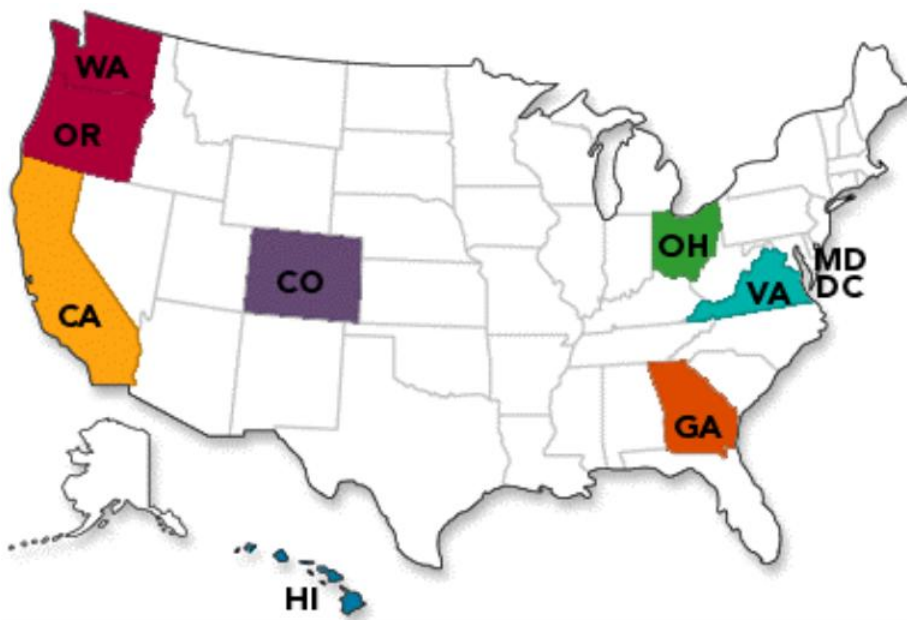
# Agenda

- **Big Data and Health Care**
  - Health Care Current State
  - Definition and Evolution of Big Data
  - Nature of Health Care Data
- **From Data Collection to Data-Driven Management**
  - Electronic Medical Record History
  - REALLY Big Data – AALIM Decision Support
  - Big Data, Big Q, Big Quality
  - Data-Driven Management – OpQ
  - Data-Driven Care – Registries
- **Health Data – The Next Frontier**
  - Predictive Modeling and Diagnostic Support
  - Personalized Medicine – Genomics

# Big Data and Health Care

# Kaiser Permanente – Nation's Largest Nonprofit Health Plan

Eight regions serving nine states and the District of Columbia



- 9 million members
- 17,000 physicians
- 176,000 employees
- 37 hospitals
- 611 medical office buildings
- 3,000 clinical research studies
- EMR – Largest, most advanced deployment
- Personal Health Record – Nearly 6 million members signed up

# Health Care Current State

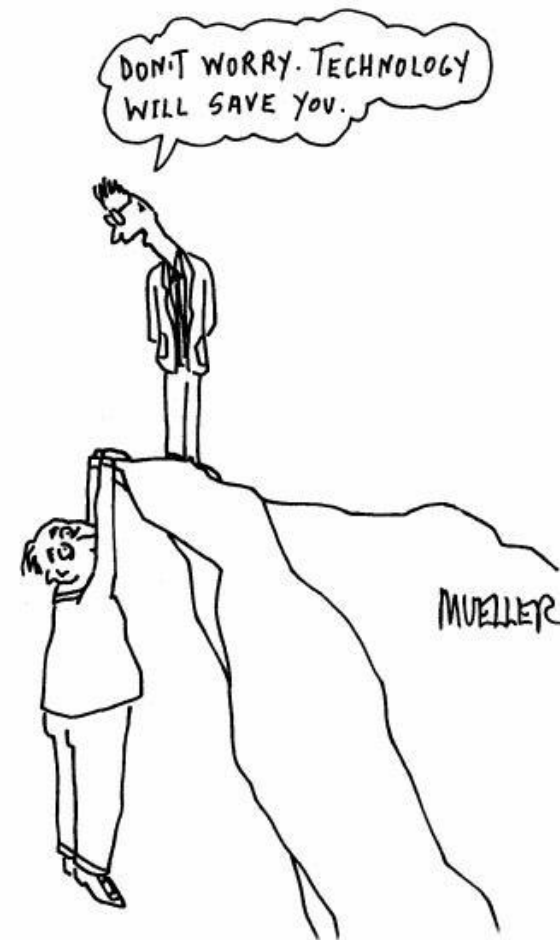
- Costs continue to soar
  - \$2.8 TRILLION
  - 18%
  - \$9,000 / year
- More than 48 million Americans are uninsured or under-insured
- Health Care Reform and new Exchanges are having a huge impact on the health care market



© Getty Images

# Current Approaches

- Automate best practices\*
  - Integrate health care across silos\*
  - Move care to less expensive venues\*
  - Move care to lower-tier workers\*
  - Engage and empower consumers\*
  - Create new financial and behavioral incentives\*
  - Create technology-enabled, data-driven care delivery models\*
- \* Massive amounts of data are generated, managed, analyzed, and used**

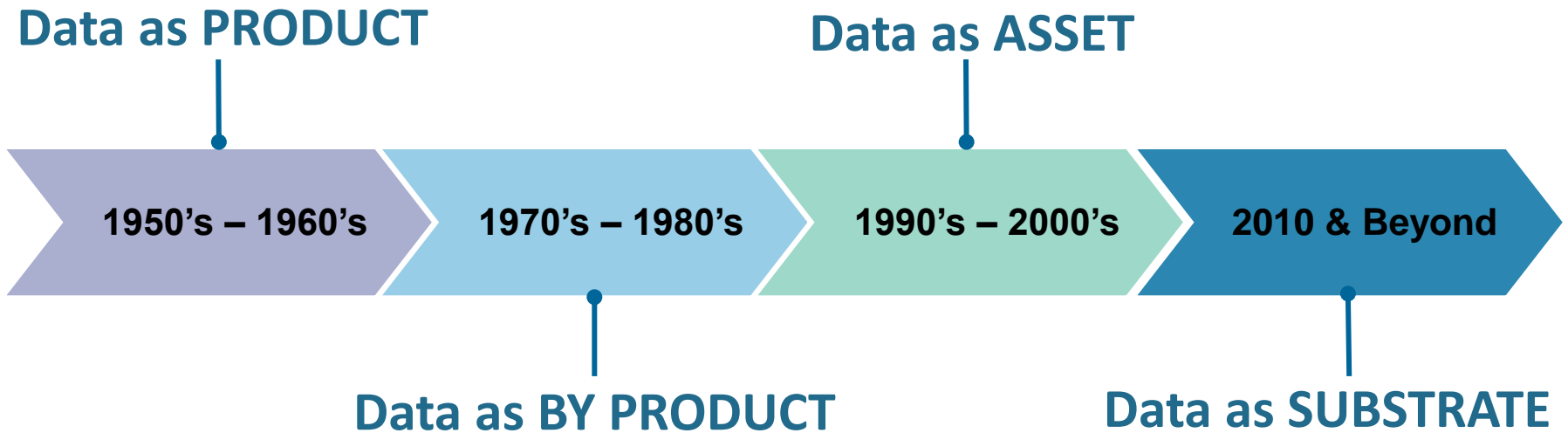


# Definition of Big Data

- Volume
- Velocity
- Variety
- Value
- Veracity (the extra “V”)



# Evolution of Data



## What Big Data is really about...

- Data ⇔ Information
- Information ⇔ Improved Decision-making



# Nature of Health Care Data

## 20% Structured

- Traditional business intelligence analytics



## 80% Unstructured

- Clinical notes, monitoring data, imaging, surveys, email, phone calls, photos, videos, location, context
- Highly complex – Multiple regulated silos, lexicons, differences in accuracy, integrity, availability, reliability, usability

# Potential Uses of Big Data in Health Care



© eHow



© ELQ



© SAIC



© fotosearch

## Care Delivery

- Decision support
- Real-time monitoring in hospital and home
- Personalized medicine
- Comparative effectiveness research
- Workflow optimization
- Cost and quality analytics
- Predictive analytics

## Operations

- Business analytics
- Marketing and supply chain analytics
- Fraud and breach detection
- Payment and pricing models
- Health economics research

## Public Health

- National and regional patient registries
- Biosurveillance
- Preventive health analytics
- Data mining for new multimodal approaches to major public health issues

## Research

- Biomedical research literature
- Clinical trial analytics
- Predictive modeling for drugs and devices
- Genomics, biodata
- Analysis of disease patterns to plan future R&D investments

# Big Data Helping Hospitals

- Hospitals are using big data to improve care and prepare for industry shifts
- To succeed, hospitals must bring doctors on board
  - Need to track whether physicians are meeting performance goals
- New tools, like Crimson, allow hospitals to track:
  - Complications
  - Readmissions
  - Cost measures for each physician



## The Daily Briefing

[Today's Daily Briefing](#) | [View Archives](#) | [Print Today's Daily Briefing](#)

### WSJ: How big data is helping hospitals get better

One health system used Crimson to save nearly \$14M, improve patient outcomes

Topics: [Business Intelligence](#), [Information Technology](#), [Big Data](#), [Advanced Analytics](#), [Quality](#), [Performance Improvement](#), [Efficiency](#)

July 12, 2013

The *Wall Street Journal's* Anna Wilde Mathews reports on why hospitals are using physician data to improve care and prepare for industry shifts, spotlighting how MemorialCare Health System used Crimson software to boost patient outcomes and save nearly \$14 million.



What is Crimson—and how can it help my hospital? From billing systems to electronic medical records, there's no shortage of useful data—but where do you start?

[This video](#) explains how Crimson Clinical Advantage is helping over 1,000 health care organizations turn the chaos of big data into big outcomes.

#### New technology helps hospitals track doctors' performance

Shifts in the U.S. health care system—partly brought on by the Affordable Care Act (ACA)—are pushing hospitals toward value-based payments and doctors toward performance-based salaries. Although these concepts have been debated before, hospital executives argue that new technology allows for better, faster tracking of individual physicians to ensure that they are meeting goals.

Source: <http://www.advisory.com/Daily-Briefing/2013/07/12/WSJ-How-big-data-is-helping-hospitals-get-better#.UeBQrXhJnok>.email

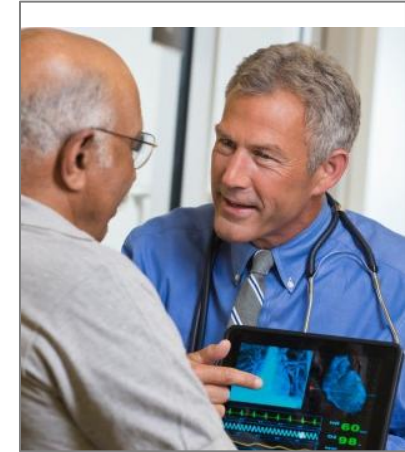
# Current Sources of Big Data



**Health Plan Members**



**Organizations**



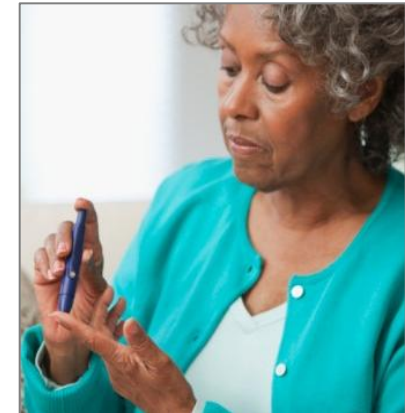
**EMR and IT Systems**



**World Wide Web**



**New Biodata and  
Biomedical Research**



**Devices and Sensors**

# **From Data Collection to Data-Driven Management**

# The Original Personal Health Record



# The Old Days...



# Where We Stand Today



- Legible
- Searchable
- Omnipresent



# 6P Vision for Medicine

**P**ersonalized

**P**redictive

**P**reventive

**P**articipatory

**P**roactive

**P**arity



**Vision**

Supported by the EMR

# Electronic Medical Record – TODAY

Patient History / Longitudinal Record

```
graph TD; A[Patient History / Longitudinal Record] --> B[Structured data]; B --> C[Guide to Best Treatment];
```

Structured data

Guide to Best Treatment

# Electronic Medical Record – TOMORROW

More complete data capture

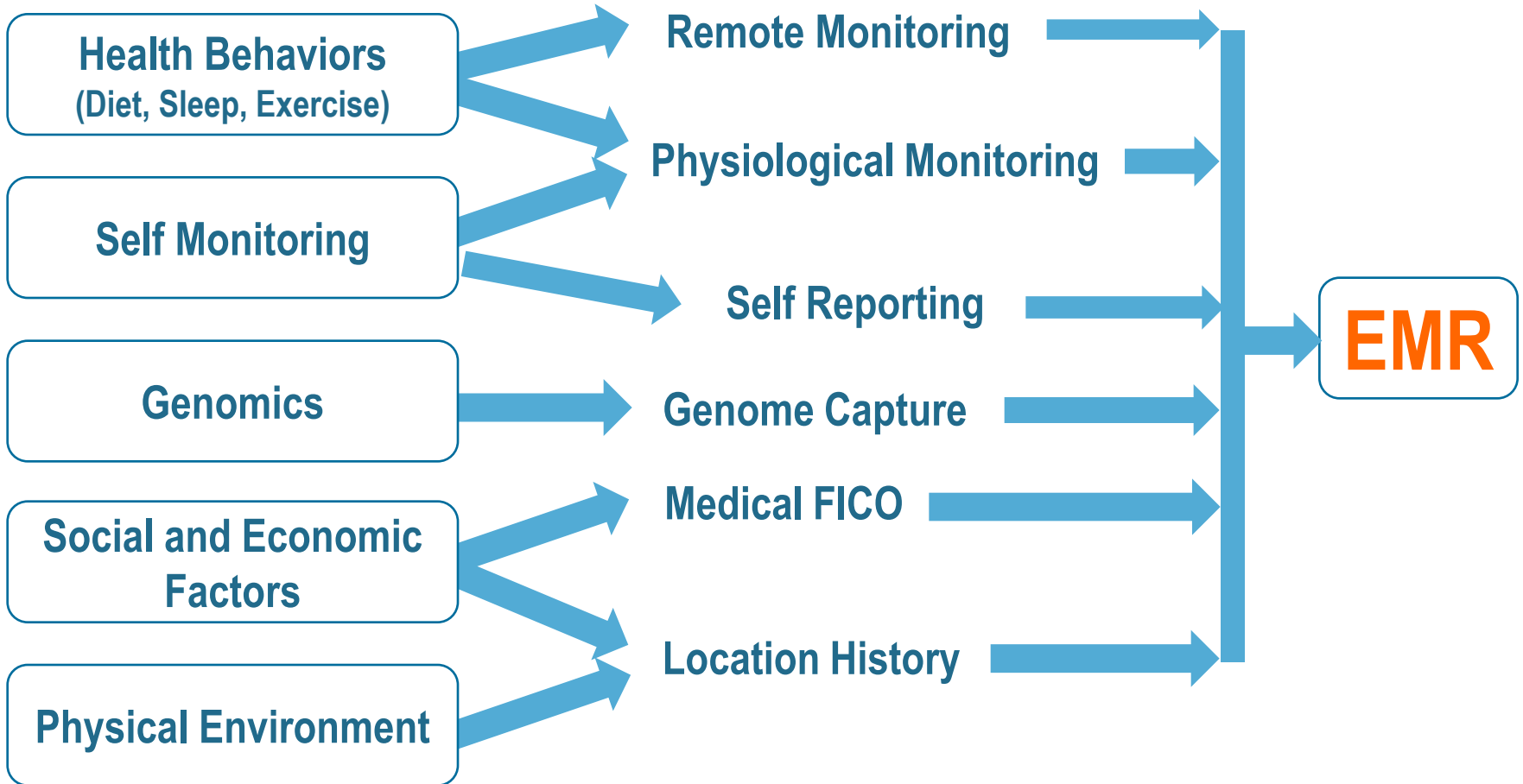
```
graph TD; A[More complete data capture] --> B[Global data access]; B --> C[Big data analytics]; C --> D[Real-time decision support];
```

Global data access

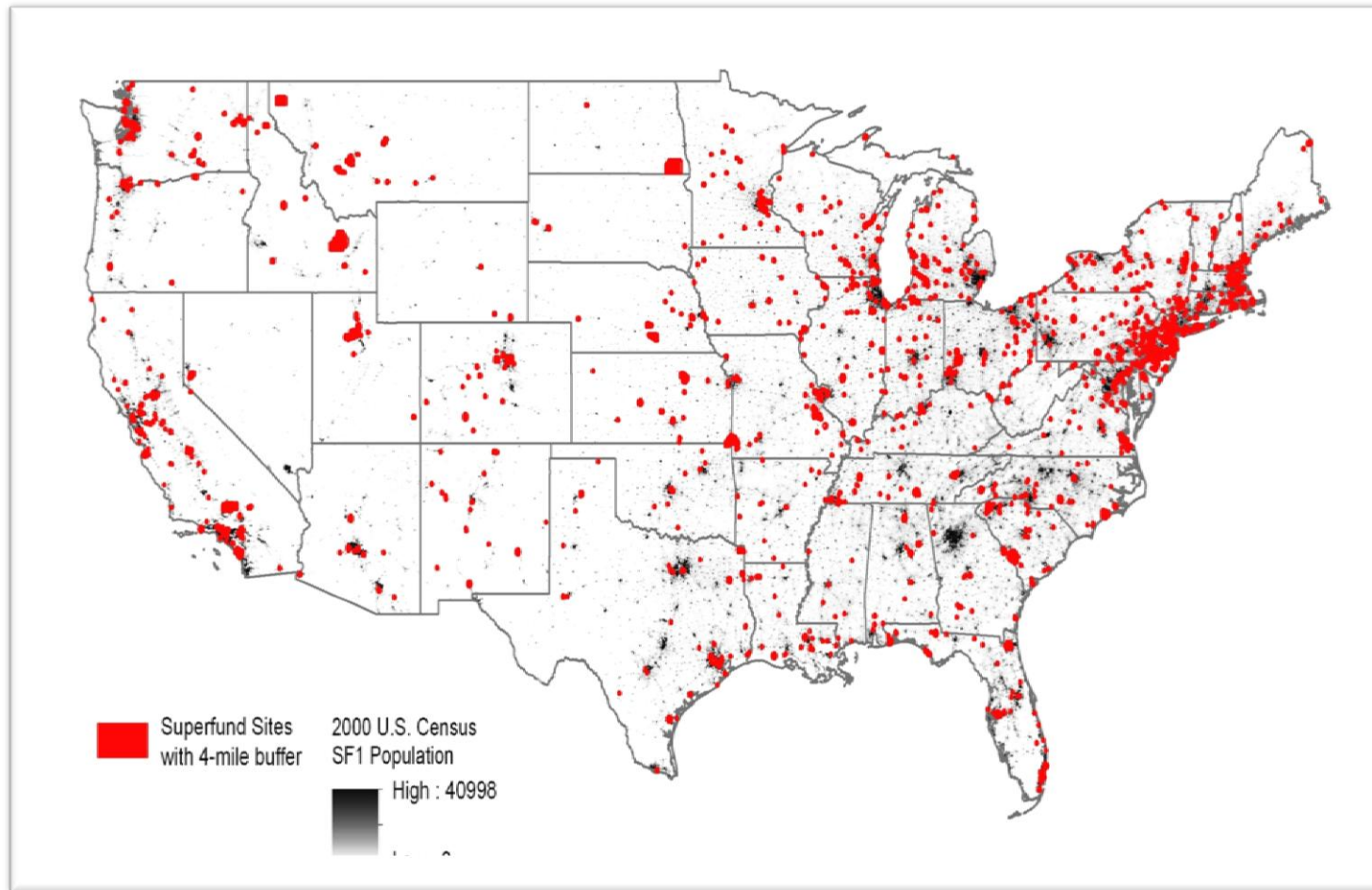
Big data analytics

Real-time decision support

# More Complete Data Capture

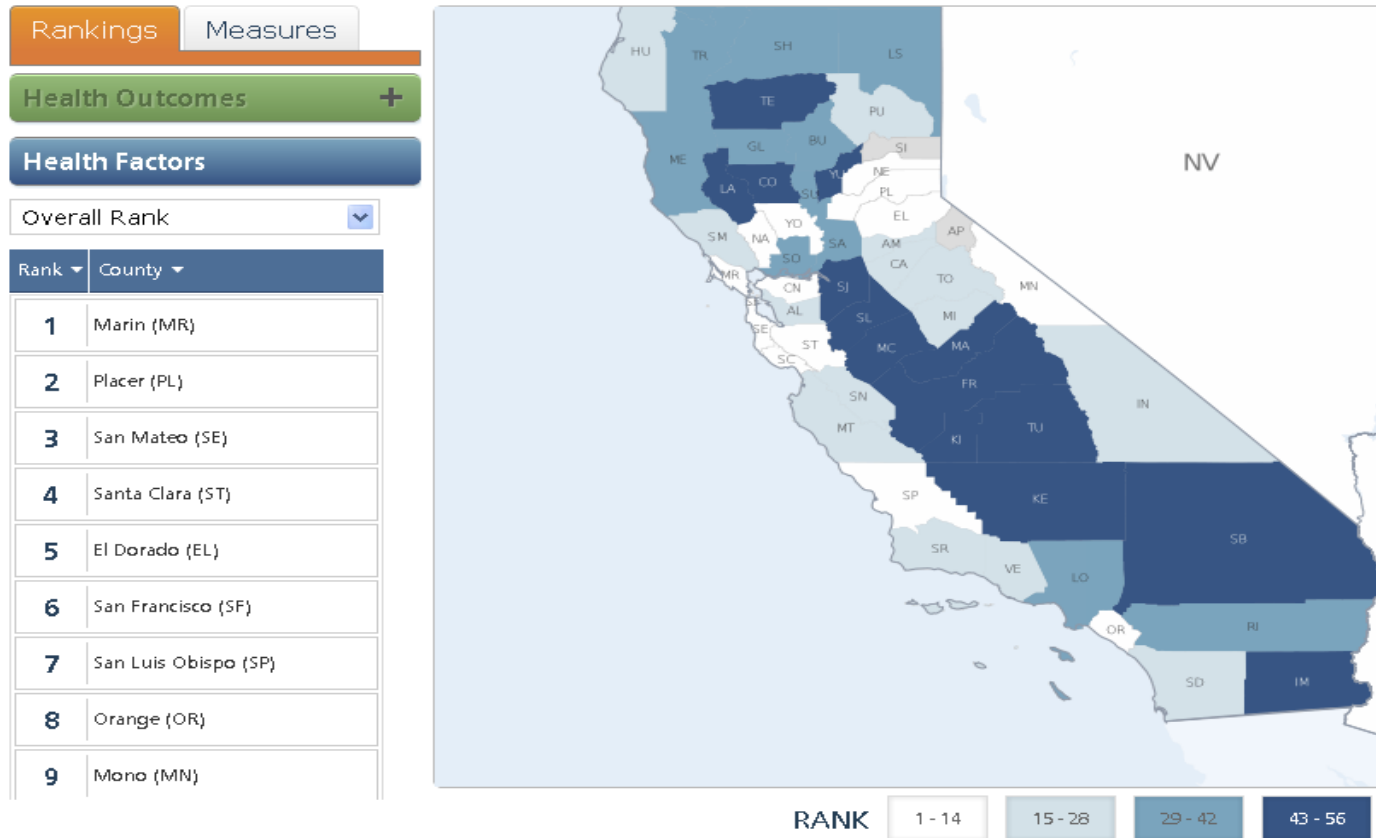


# Location History – Superfund Locations



Source: EPA

# Location History – County Health Rankings



Source: <http://www.countyhealthrankings.org/>

# Location History – Food Deserts

Home Enter Locator About the Locator Documentation Download the Data More Maps ERS Website

## Food Desert *Locator*

Get a spatial overview of low-income neighborhoods with high concentrations of people who are far from a grocery store.

*Map food deserts and view census tract-level statistics on population groups with low access to healthy food.*

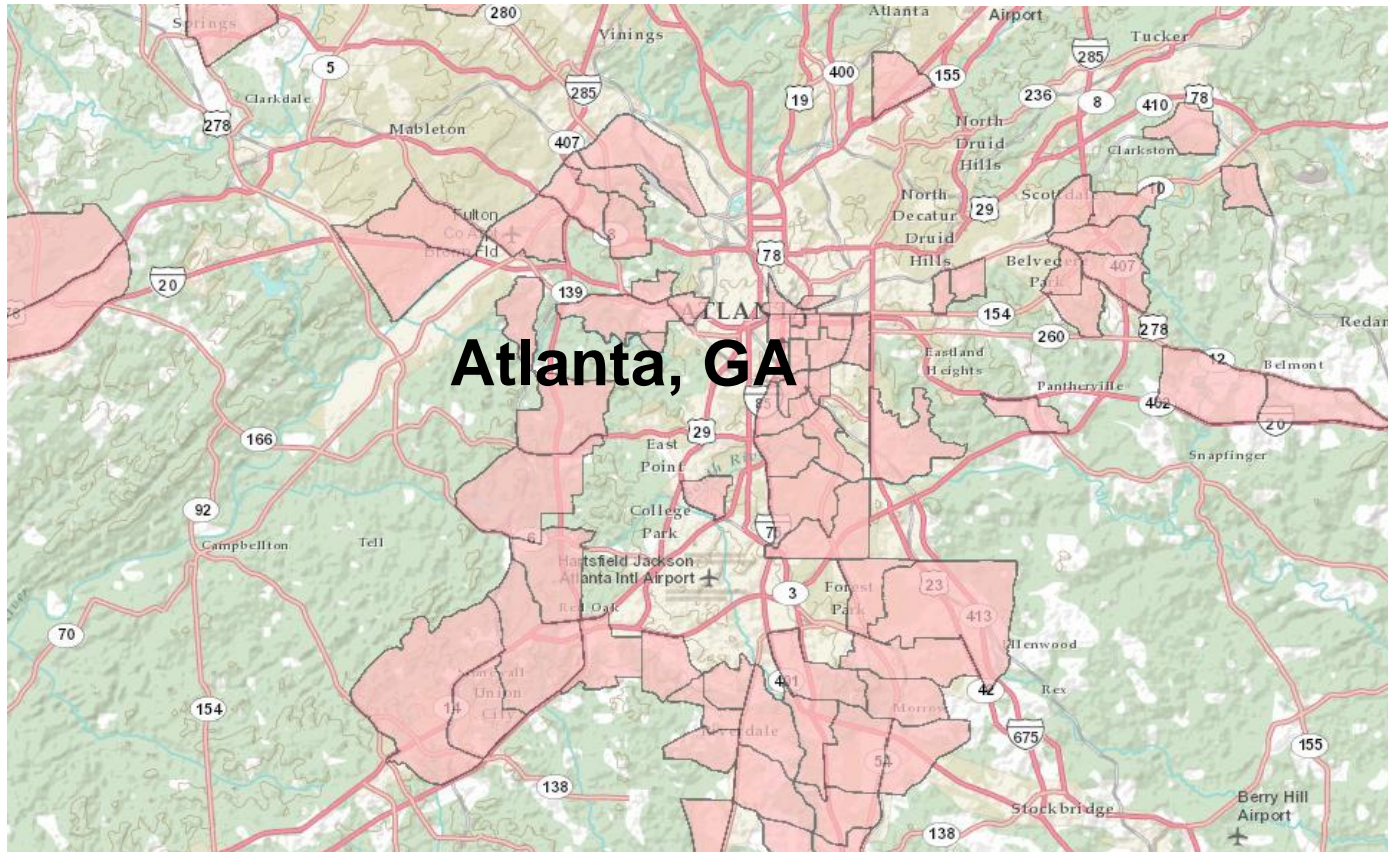
[Enter Locator](#)

U.S. Department of Agriculture  
Economic Research Service

You may also be interested in ERS's [Food Environment Atlas](#) [Enter](#)

Source: U.S. Department of Agriculture

# Location History – Food Deserts



Source: U.S. Department of Agriculture



**REALLY Big Data**

# IBM's AALIM Decision Support System

## Advanced Analytics for Information Management

- Extracts, analyzes and correlates information across patient records
- Allows statistical validation of diagnosis by gathering evidence from similar patient data sets
- Summarizes the outcomes, history and diagnosis for similar patient data
- Brings the physician the right information at the right time leading to more informed decision-making



Source: <http://research.ibm.com/index.shtml>

### Key idea behind using AALIM

By finding similar patients from pre-diagnosed data sets, diagnosis – and hence treatment and outcomes – can be inferred for a new patient

# AALIM Focus: Cardiac data

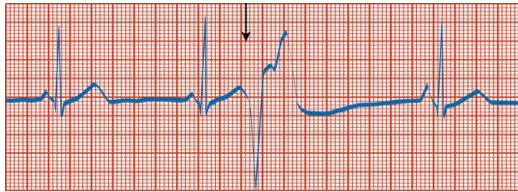
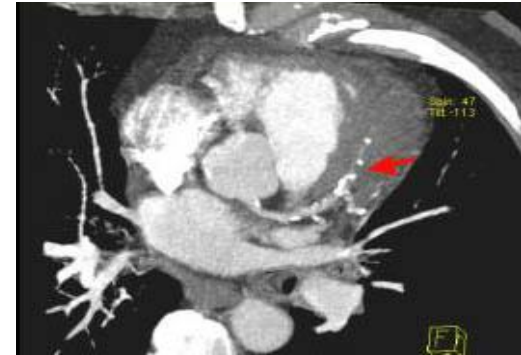


Figure 17-36 R-on-T premature ventricular contraction. (From Huff J. ECG Workout [4th Ed]. p 195. Philadelphia, Lippincott Williams & Wilkins, 2002.)  
Copyright © 2009 Lippincott Williams & Wilkins. Instructor's Resource CD-ROM to Accompany Critical Care Nursing: A Holistic Approach, eighth edition.

```
<section>
<code code="8709-8" codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC"/>
<title>Skin Edem</title>
<text>Erythematous rash, palmar surface, left index finger.
</renderMultiMedia referencedObject="MM2"/>
</text>
<entry>
<observation>
<code code="10676001" codeSystem="2.16.840.1.113883.6.96"
codeSystemName="SNOMED CT" displayName="Skin finding"/>
<value xsi:type="CD" code="271607003"
codeSystem="2.16.840.1.113883.6.96"
codeSystemName="SNOMED CT" displayName="Rash"/>
<targetSiteCode code="4695004"
codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT"
displayName="Skin of palmar surface of index finger">
<qualifier>
<name code="78615007" codeSystem="2.16.840.1.113883.6.96"
codeSystemName="SNOMED CT" displayName="with laterality"/>
<value code="7771000" codeSystem="2.16.840.1.113883.6.96"
codeSystemName="SNOMED CT" displayName="left"/>
</qualifier>
</targetSiteCode>
<entryRelationship typeCode="SPRT">
<RegionOfInterest name="MM2"
id root="10.23.4567.4489"/>
<code code="EJJPSE" value="31372444" value>
<entryRelationship typeCode="SUBJ">
<observation data type="ED" mediaType="image/jpeg">
<value xsi:type="ED" mediaType="image/jpeg">
<reference value="leftand.jpeg"/>
</value>
</observation>
</entryRelationship>
</RegionOfInterest>
</entryRelationship>
</observation>
</entry>
</section>
```

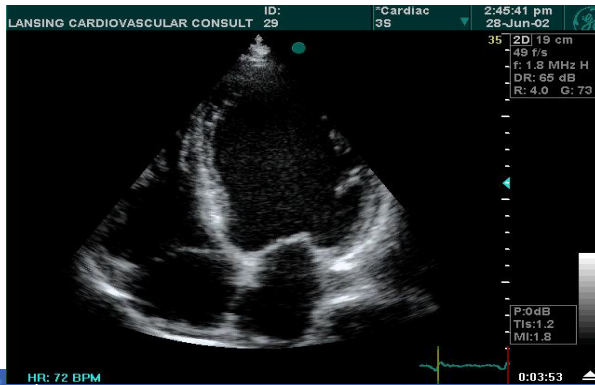


EKG Images/Time Series

HL7, CDA, XDS  
UMLS, MESH

Cardiac CT  
images (DICOM)

Cardiac  
echo  
video &  
MRI



Patient  
demographic  
information from  
EMR



Audio

# AALIM – Kaiser Permanente’s Cardiac Project

## ■ Subjects:

- Evaluated 300,000 patient records
- Used 1,500 for analysis due to complete data sets
- User testing occurs on roughly 20 new patients a day at the KP San Francisco Cath lab

## ■ Data sources:

(ICD9) OSCR

List of ongoing problems

Medications prescribed

Free text

“Signal” traces of ECG reports

Dictated surgical reports from catharization procedures

Diagnosis

Patient demographics

Vitals

ECG reports

Echocardiogram reports

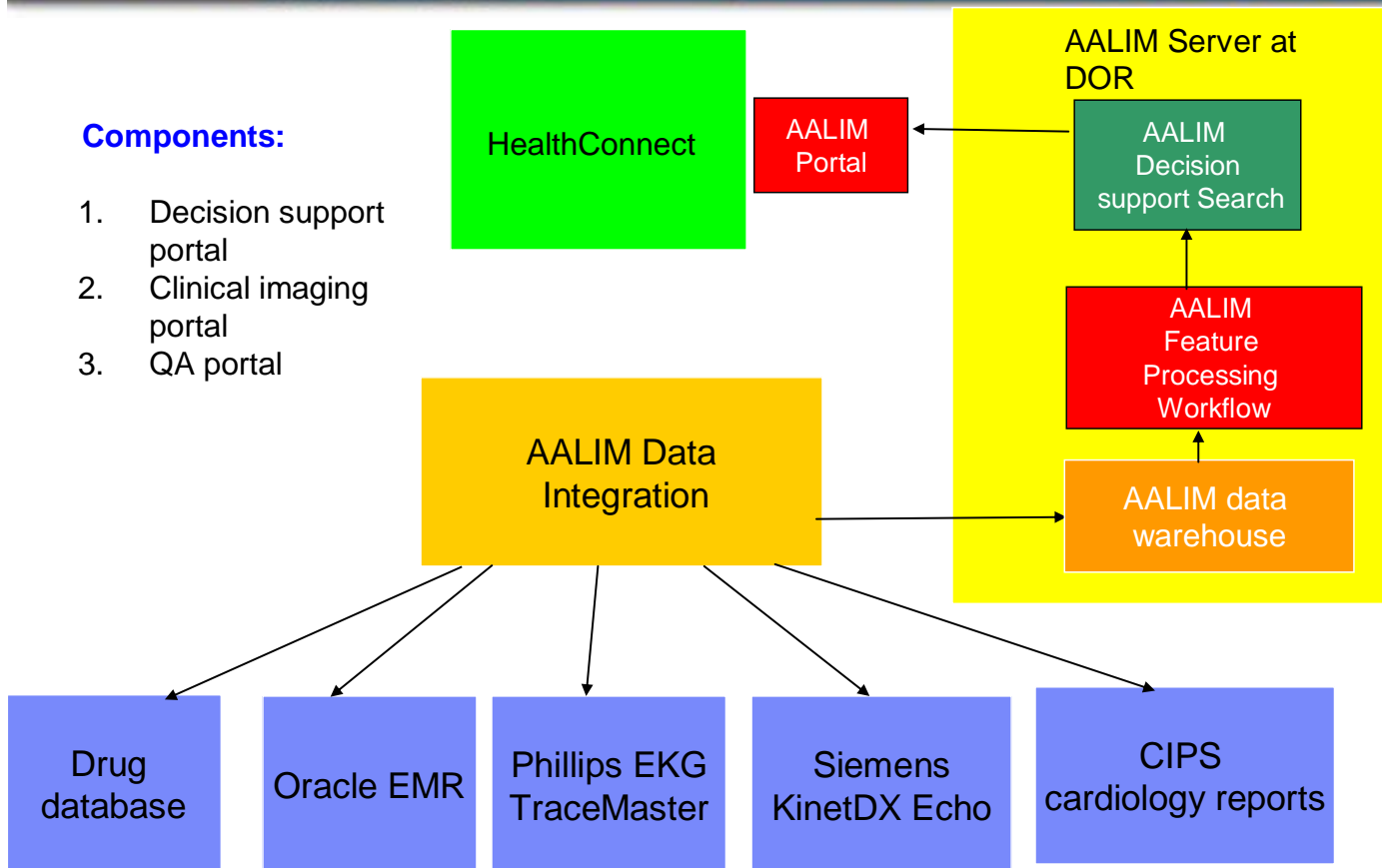
Streamed video from echocardiogram

# KP HealthConnect and AALIM Interface



## Components:

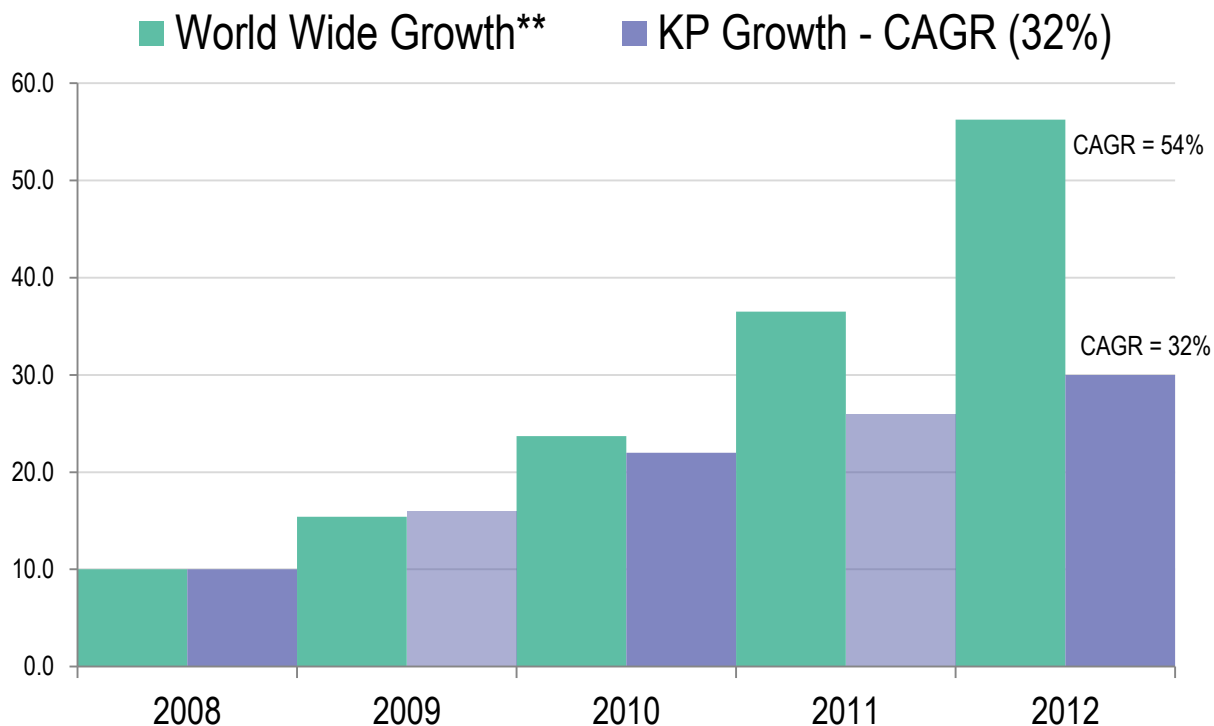
1. Decision support portal
2. Clinical imaging portal
3. QA portal



# What We Found

	Mitral Regurgitation Identification	Mitral Stenosis Identification
AALIM	95% accurate	92% accurate
OSCR	50% – 75% accurate	45% -- 75% accurate

# Demand for Storage is Increasing within Kaiser Permanente and World Wide



\*\* Source: IBM for World Wide Growth  
\*\* Growth excludes Teradata

International Data Corp. projects a **61.4%** five-year Compounded Annual Growth Rate (CAGR) through **2015** for world-wide storage.

**Big Data, Big Q, Big Quality**



# Kaiser Permanente Big Q = Big Data

**Kaiser Permanente's Big Q dashboard is a comprehensive and integrated view of our Quality and Service performance**

*It's all about improving the quality of the care we provide to our patients and members*

# Big Role for Big Q

- **Consolidates national strategic reporting into a single application**
  - Key source for quarterly reporting on Quality and Service measures
- **Displays “big dot” view of Quality and Service measures**
  - Top-level of overall performance
  - Detail-level by specific Region or Hospital
  - Trend and comparative variation over time for each measure
  - Comparative metrics for benchmarking across organizations

# We describe each area of the Big Q as “Big Dots”



# Clinical Risk Management

## Key Enablers

Culture

Leadership

Capacity

Operational Links

Data/Measures

## Outcome

**No Preventable Harm**

**All Possible Repair**

*We exist to support the organization in its pursuit of safe, reliable, and patient-centered care. We are committed to collaborative problem solving and continuous performance improvement toward an end-state of no preventable harm for our patients. If harm does occur, in order to achieve all possible repair for the patient/family, provider/staff and organization we will be honest with our patients, open with our colleagues and ourselves, and able to handle such occurrences with sympathy and empathy*

## Key Drivers

Maintaining Relationships

Preventing Harm

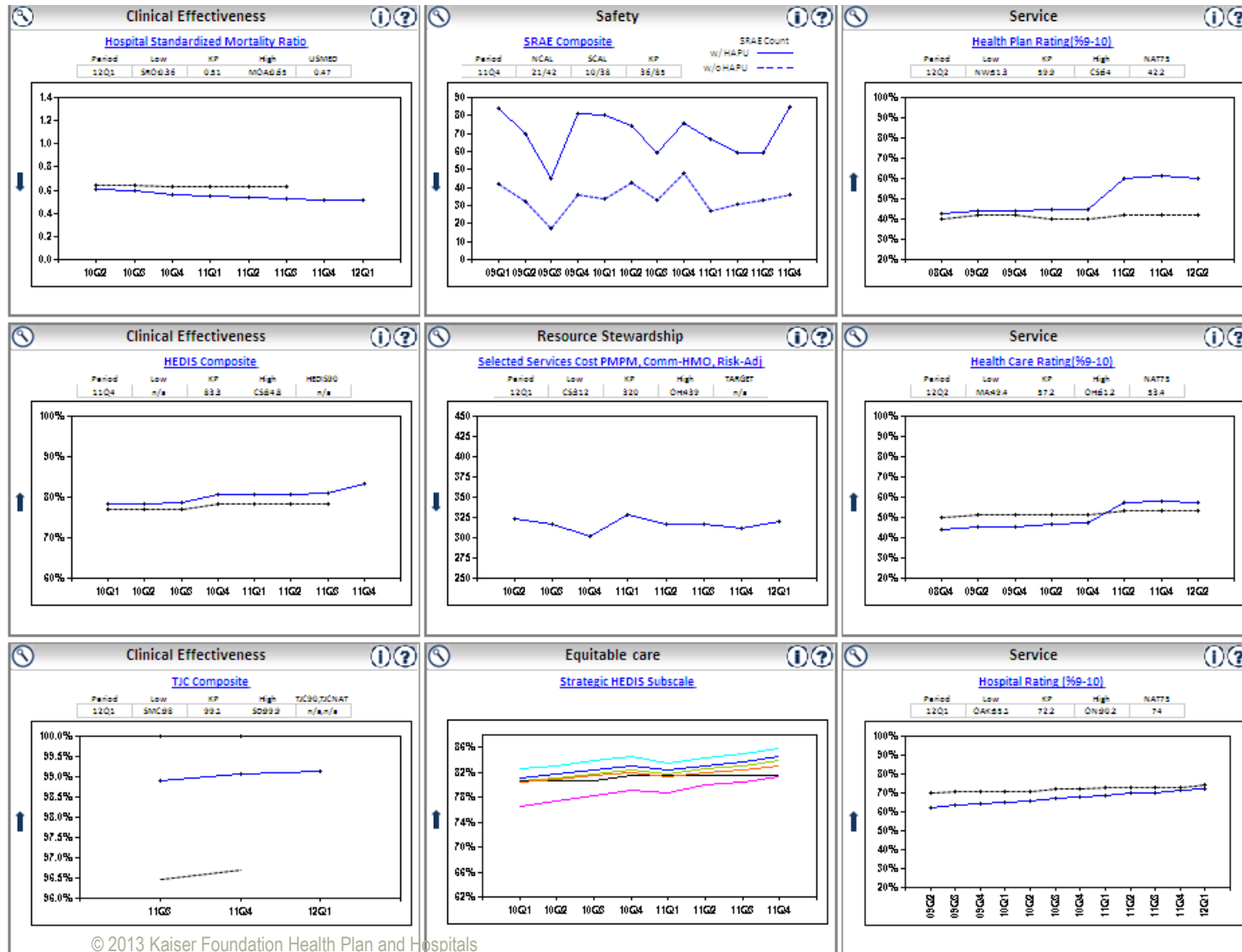
Responding to Injury

## Focus Areas & Initiatives

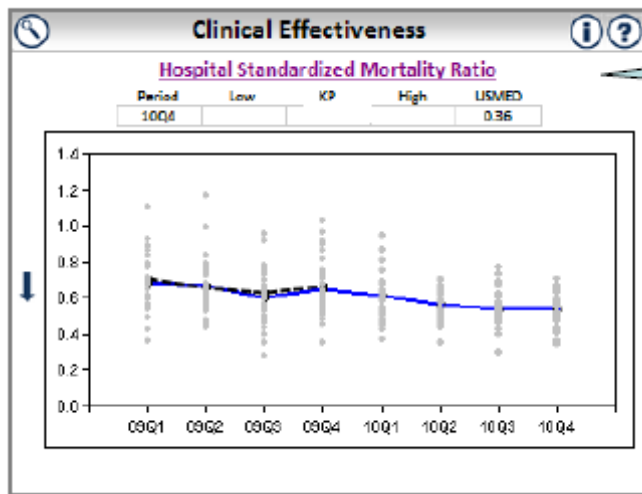
- Patient and Family Centered Care
- Health Literacy
- Service Quality
- Membership Services
- Safety Culture
- Perinatal Patient Safety (PPSP)
- Highly Reliable Surgical Teams (HRST)
- Diagnostic Reliability (DRII)
- Emergency Medicine Risk Initiative (EMRI)
- Simulation Based Education
- High Alert Medication (HAMP)/KP MedRite
- Trigger Tools
- Infection Prevention and Control
- Mortality Reduction
- KP HealthConnect
- Just Culture/Responsible Reporting
- Healthcare Ombudsman/Mediator(HCOM)
- CUAO/SMT
- Early Resolution
- Risk Management Education
- ASPIRE-Learning from Events
- Complaint Management

# The Big Q Dashboard Houses Our Data

## It's How We Measure Our Success



# Big Q – Two Display Types

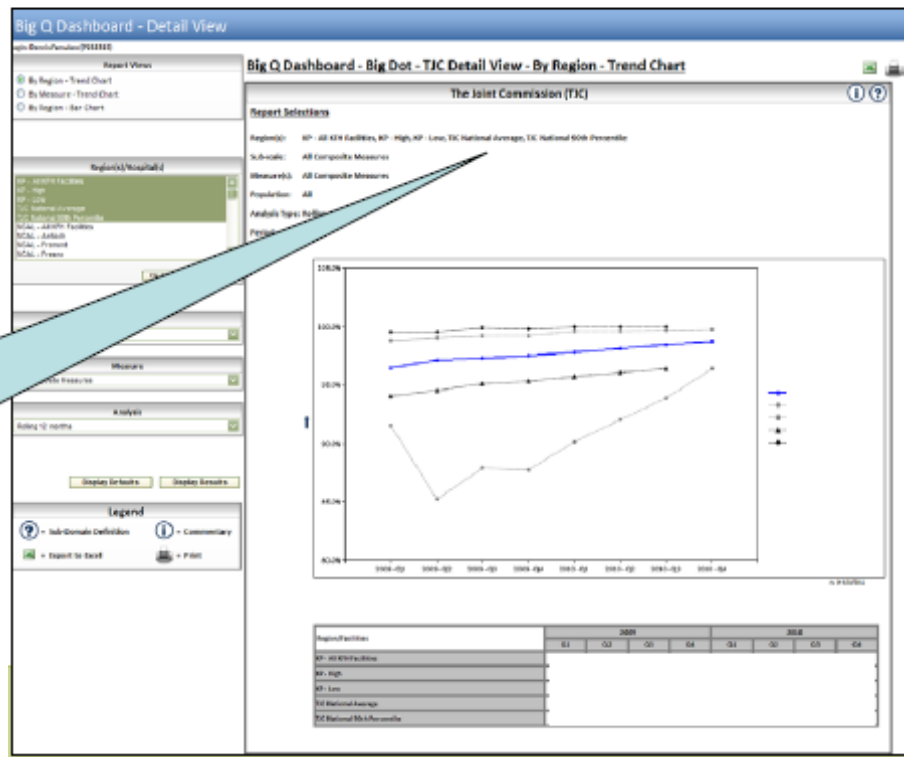


**Top Level:**

- ✓ Aggregate Reporting
- ✓ Current Performance
- ✓ Trend over time
- ✓ Variation
- ✓ Benchmark Comparison

**Detail Level:**

- ✓ Aggregate to Individual Measures
- ✓ User-Defined Reports via Parameters
- ✓ Region & Med Center Levels
- ✓ Benchmark Comparisons



# Data-Driven Management

# OpQ Project – Goals

OpQ is a web-based application coalescing data from Chronicles, Clarity, and data warehouses to provide outpatient front-line clinician leads with near real-time data impacting multiple aspects of operations in one dashboard.

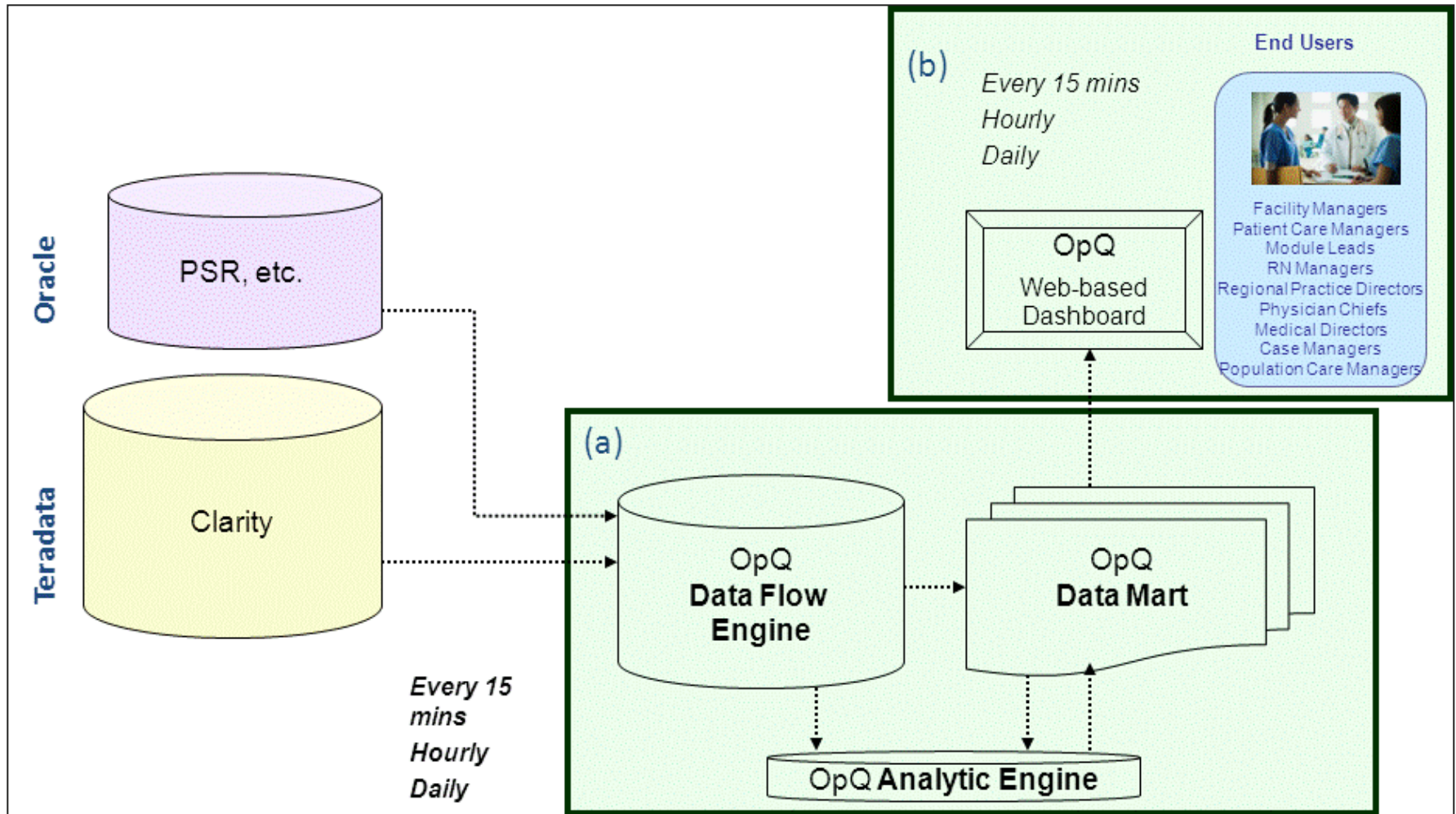
- **Operational Efficiency**
  - Proactively schedule staff
- **Quality of Care**
  - Timely, reliable
  - Documentation of quality metrics
- **Service**
  - Maximize appointment utilization
  - Decrease response time to patient calls and emails
- **Professional Satisfaction**
  - Empower frontline to drive improvement





# OpQ – Up-to-date Organizational Metrics

Provided by proprietary technology & monthly data flows



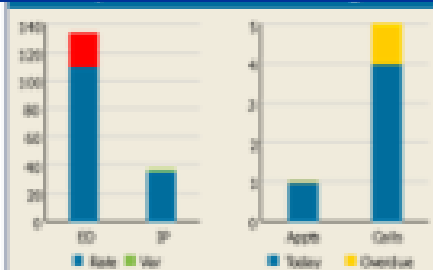
# OpQ – Data Available from 3 Views

## Primary Care, Specialty Care, and Appointments

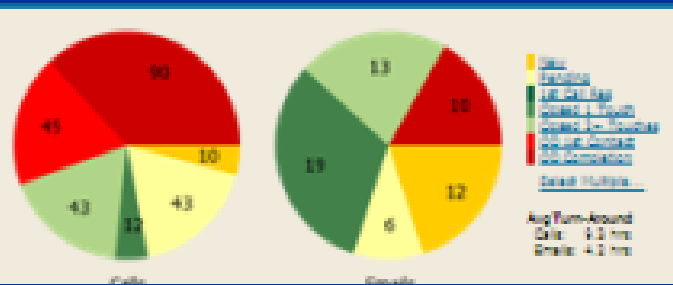
OpQ PC      OpQ SC      OpQ APPT

OpQ PC      OpQ SC      OpQ APPT

### ED/IP Admissions and Discharges



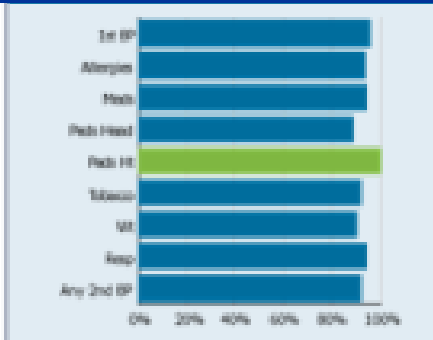
### Patient Initiated Calls and Emails



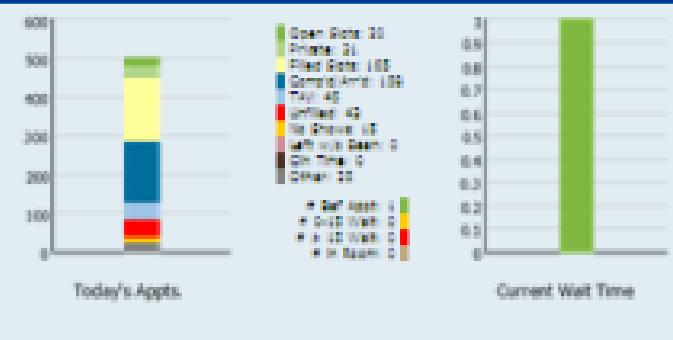
### Selected Specialty Referrals



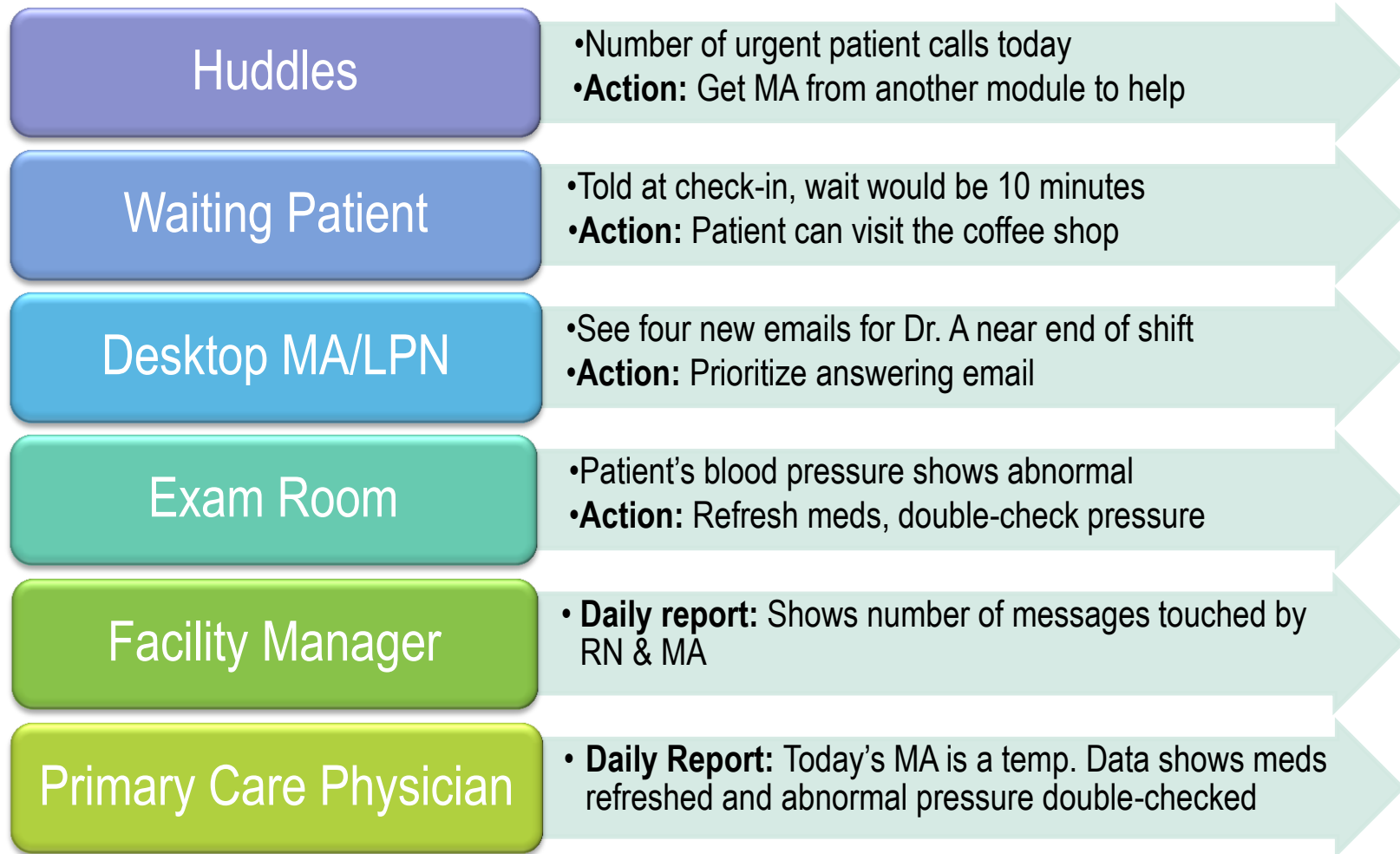
### Clinical Quality Service



### Office Visits

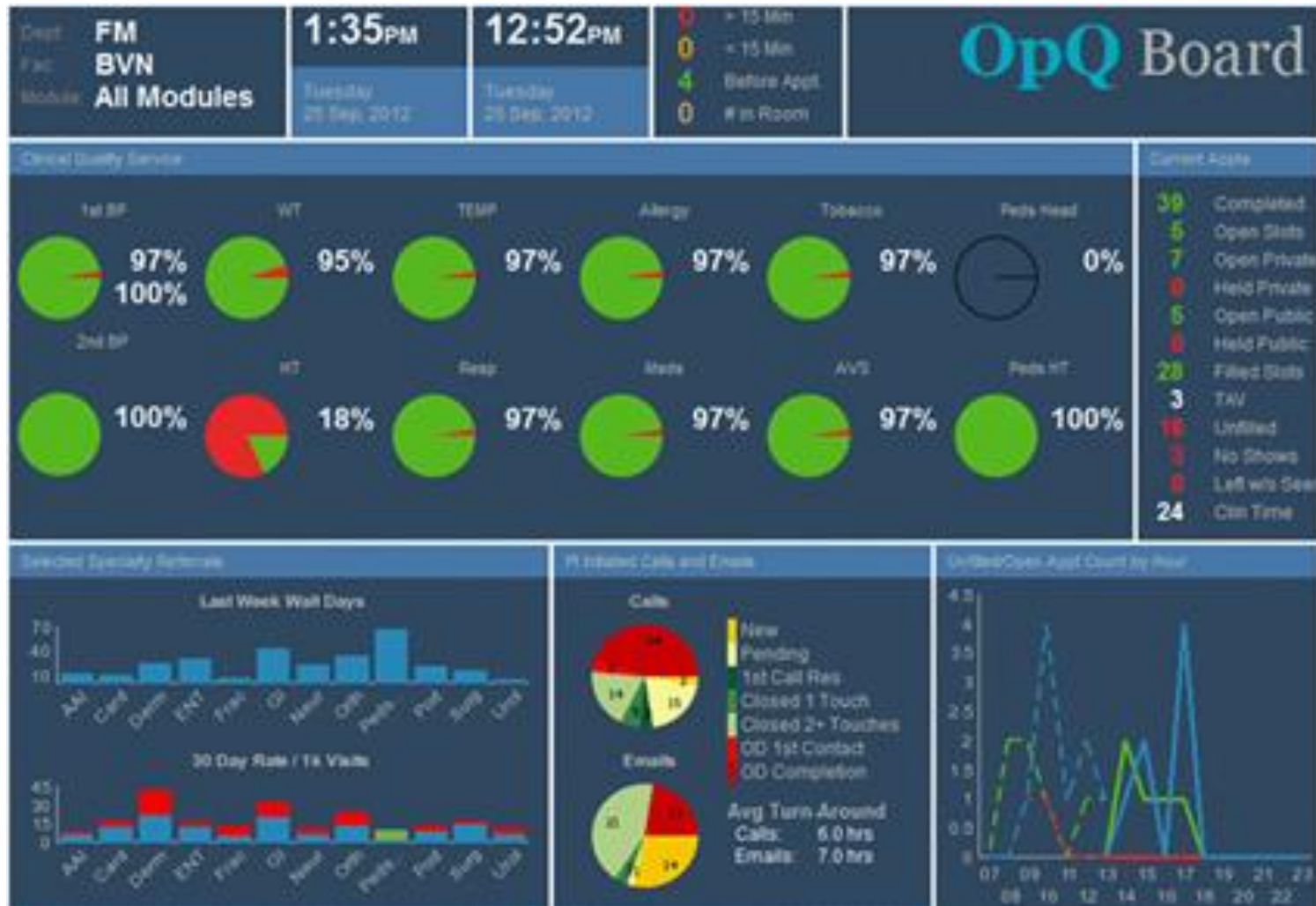


# OpQ – Data from 3 Views Used to Take Action in 6 Key Areas



# OpQ – Data-Driven Management

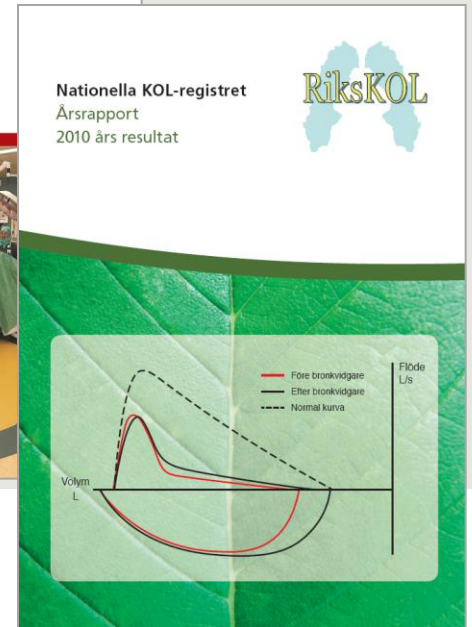
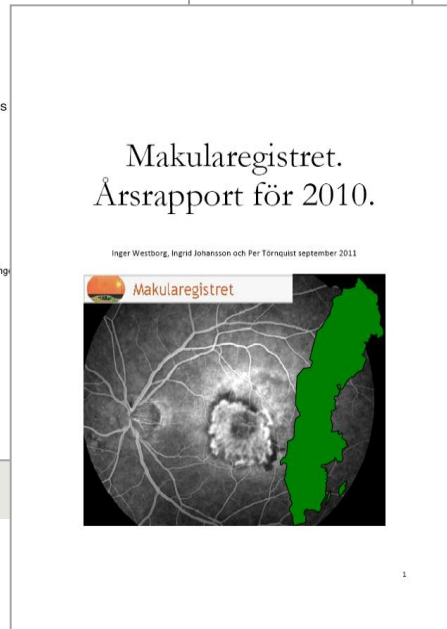
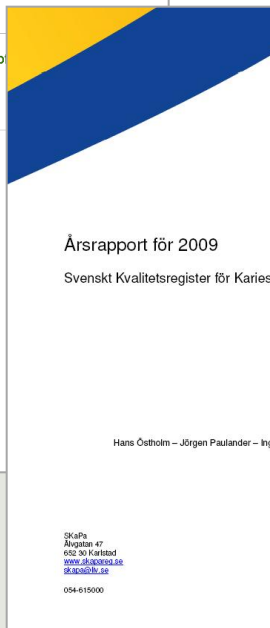
Staff can view clinical quality measures as they happen on an LCD screen in clinical work areas



# Data-Driven Care

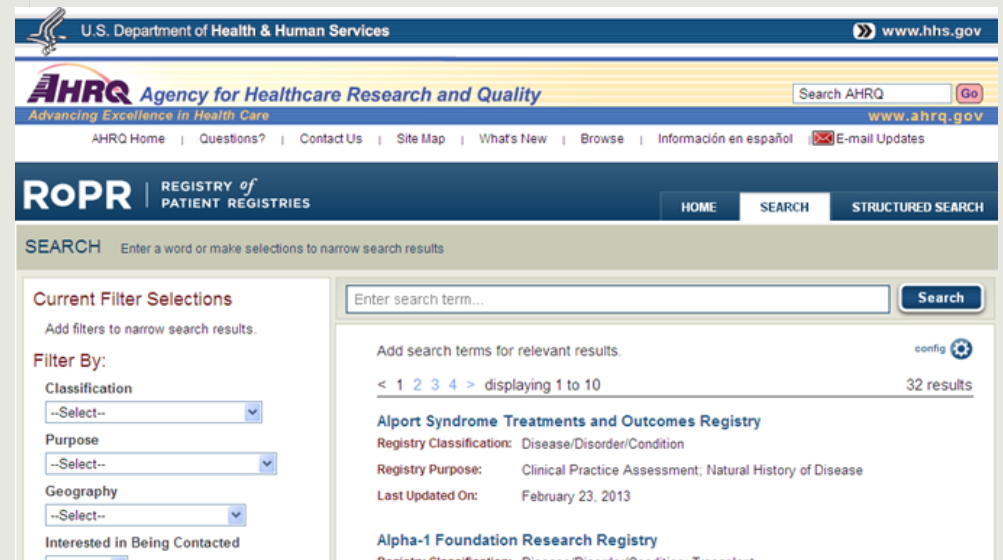
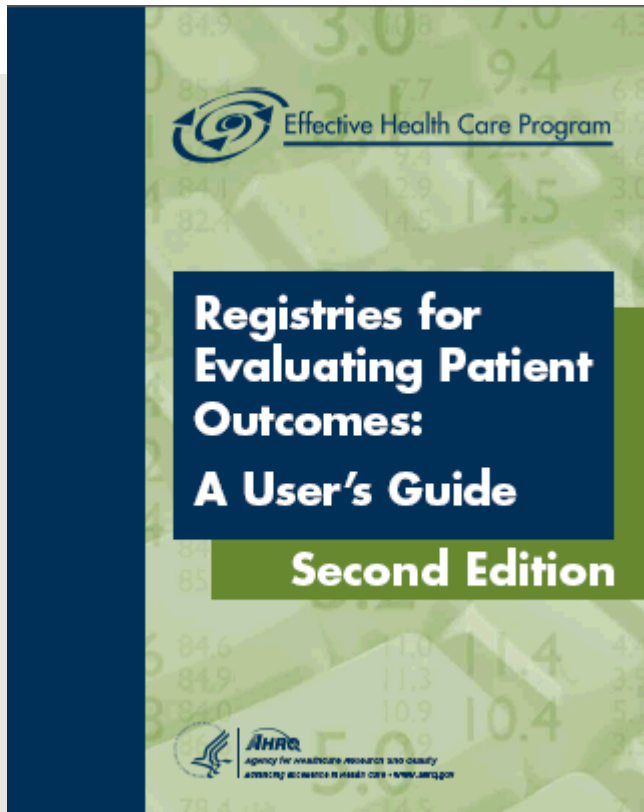
# Registries – A Multitude of Applications

## Swedish Examples



Source: <http://www.kvalitetsregister.se>

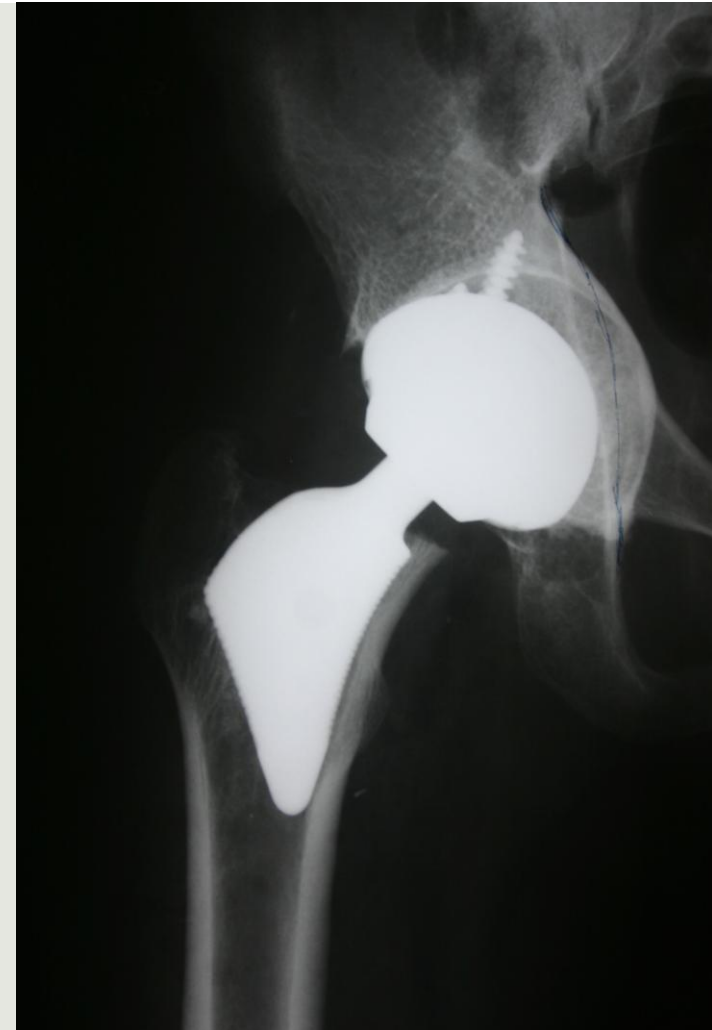
# AHRQ Support and Guide to Patient Registries



Source: [https://patientregistry.ahrq.gov/search#search/?init=true&\\_suid=13621737551940019672112306579947](https://patientregistry.ahrq.gov/search#search/?init=true&_suid=13621737551940019672112306579947)

# Implant Registries – Improving Quality of Care

- Collect information on
  - Demographics
  - Implant models
  - Medical procedures
  - Health outcomes
- How do registries help?
  - **Track success of implants**
    - Help health teams determine the right implant
  - **Improve quality of care**
    - Registry data helps predict risks for complications
  - **Increase patient safety**
    - In rare cases of recalls, registry allows providers to identify affected members





# Kaiser Permanente Total Joint Registry

## “Patients Like Me”

Orthopedics  
Total Joint Replacement

Reports

- Specifications
- Annual Reports
- Dynamic Registry Report
- Infection Control
- Participation Reports
- Risk Calculators**
- TJR Total Hip
- TJR Total Knee

Sample Forms

Champions

ACLR

Spine

Hip Fracture

**Cardiology**

**Cardiac Surgery**

### TJR Total Knee

**Age:**  years  months

**Gender:**  Male  Female

**Height:**  feet  inches

**Weight:**  pounds

**Diabetes:**  No  Yes

**Osteoarthritis:**  No  Yes

**Inflammatory Arthritis:**  No  Yes

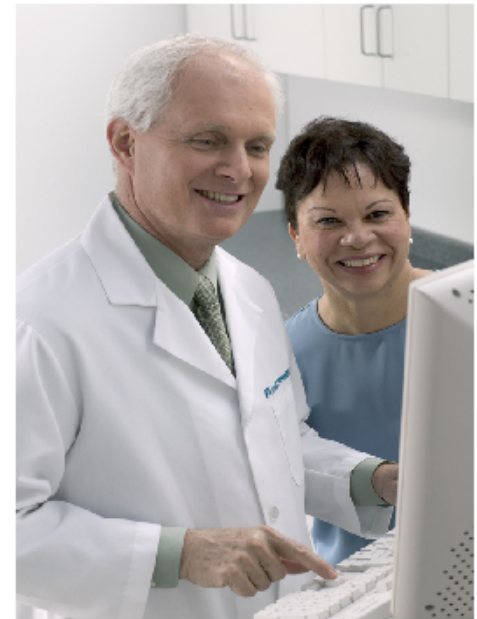
**Post Traumatic Arthritis:**  No  Yes

**Rheumatoid Arthritis:**  No  Yes

**Osteonecrosis:**  No  Yes

Your risk of a revision is (within 5 years):

Javascript and ActiveX controls are required to use calculators.



# Impact of Registry Data on Practice

- **Decreases in:**
  - unicompartmental knees
  - uncemented knees
  - mobile bearing knees
  - minimally invasive surgeries
  - smaller femoral hip components
  - hip resurfacing procedures
- **Identifies patients at risk for:**
  - postoperative infections
  - second surgeries
  - hospital readmissions
  - other complications
- **Helps in evaluation** of manufacturers' marketing claims and response to product recalls



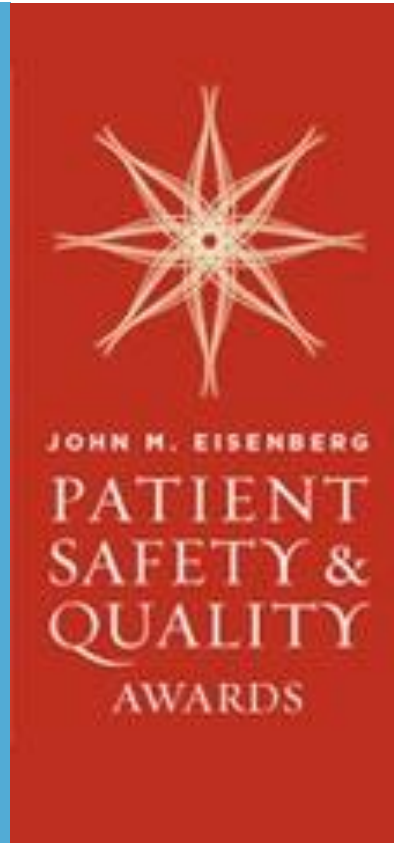
Source: <http://www.thepermanentejournal.org/files/Winter2013/ElectronicHealthRecords.pdf>

# Kaiser Permanente's Implant Registry

2012 Eisenberg Patient Safety & Quality Award Winner

“The pioneering innovations of their implant registries have shown unsurpassed and proven benefits for patient safety, quality, outcomes and cost effectiveness.”

— National Quality Forum and  
The Joint Commission



# Health Data – The Next Frontier

# Predictive Modeling

Modeling Healthcare

**ARCHIMÉDES**  
Quantifying Healthcare

$$F^k(i) = \langle F(i) \rangle + \sum_{j=0}^l \sum_{i=0}^l \psi_i^k \psi_j^k = \langle F(i) \rangle + \sum_{j=0}^l \sum_{i=0}^l \psi_i^k \psi_j^k$$

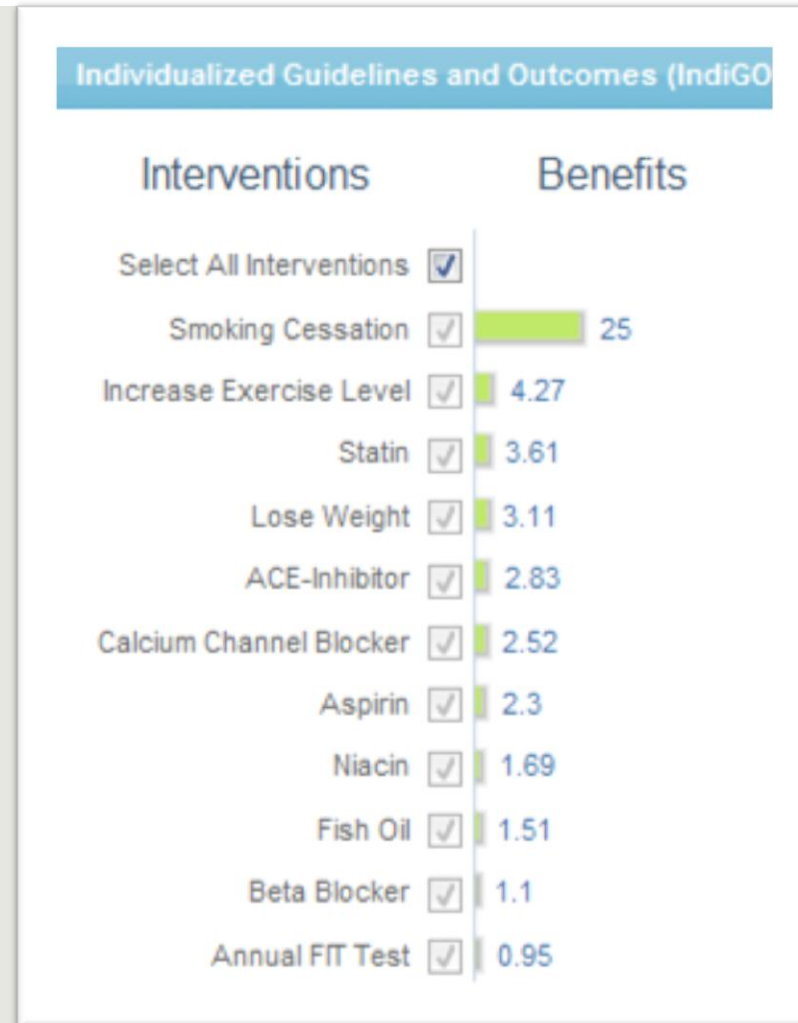
$$p_{ij}(x|\theta_1^i, \theta_2^i, \dots, \theta_N^i) = p(x < 2^{ii}(\omega) > x)$$

$$F^k(i) - \sum_{j=0}^l p_{ij}^k(i) - \sum_{j=0}^l p_{ij}^k(i)$$

$$p_{ij}(x|\theta_1^i, \theta_2^i, \dots, \theta_N^i) = p(x < 2^{ii}(\omega) > x)$$

# Predictive Modeling: Individualized Outcomes

- **IndiGo considers:**
  - Over 30 risk factors
  - Risk of high-cost preventable events
  - Effect of potential treatments
- **IndiGo can:**
  - Combine multiple outcome effects into one benefit score for each treatment
  - Identify patients who don't qualify for standard guidelines
  - Rank interventions
  - Prioritize panels of patients
  - Educate patients



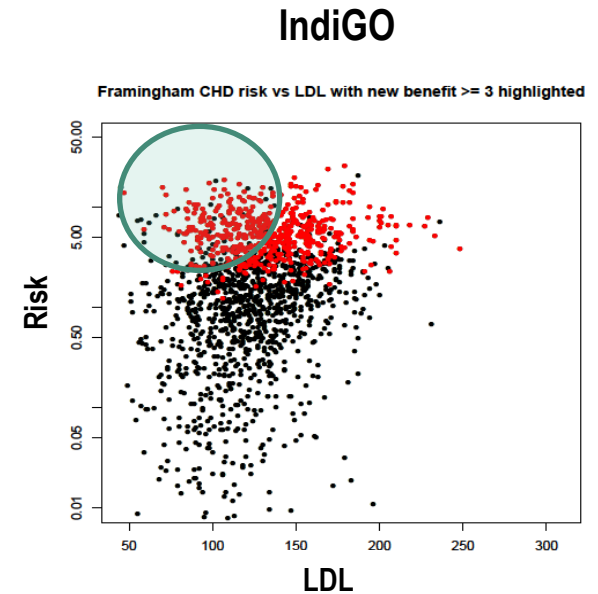
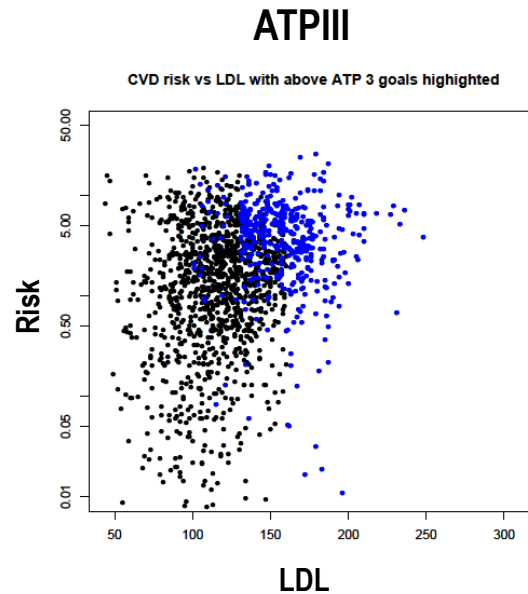
# Comparison of IndiGO and ATPIII Statin Guidelines: Who is Treated?

- **ATPIII**

recommends  
statins to people  
with high LDL  
(blue dots)

- **IndiGO**

recommends  
statins to people  
based on risk  
(red dots)



# GNS Healthcare

## Using big data analytics to discover what works in health care

- The GNS REFS™ analytics platform helps:
  - Spot causal patterns in health data
  - Reveal hidden interactions of treatments, care processes, conditions and patient characteristics that compromise safety or cause adverse events
  - Chart complex treatment or service pathways that improve health
  - Recognize which treatments or services are more effective or where gaps exist

**GNS HEALTHCARE** DISCOVERING WHAT WORKS. AND FOR WHOM.

HOME ABOUT SOLUTIONS NEWS & EVENTS CAREERS CONTACT US

### Discovering What Works in Healthcare.

GNS Healthcare is a big data analytics company that has developed a [scalable approach](#) for the discovery of what works in healthcare, and for whom. Our analytics solutions are being applied across the healthcare industry: from [pharmaceutical and biotechnology companies](#), [health plans and hospitals](#), to integrated delivery systems, Pharmacy Benefits Managers (PBM), and Accountable Care Organizations (ACOs). Whether your organization is delivering care or developing personalized therapies and diagnostics, GNS Healthcare can help you discover the knowledge you need to match patients with treatments that work.

#### GNS HEALTHCARE in the News

**AUGUST 7, 2013 | Events**  
Qualcomm Life Connect  
CEO Colin Hill will present a talk entitled "Big Data: Value to the Enterprise" at Qualcomm Life Connect on August 7, 2013 at 8:45am... [More](#)

**JULY 26, 2013 | Events**  
GNS Healthcare Co-Founder Recognized at White House Dinner  
President Obama recognized contributions of Muslim American entrepreneurs, including GNS Healthcare co-founder Iya Khalil, at an Iftar dinner at the White House Thursday night. [More](#)

**JULY 2, 2013 | Media**  
eWeek: GNS, Cambia Use Big Data to Study Medication Adherence, Health Costs  
A big data analytics platform from GNS and Cambia will spot causal patterns in health data to find ways to keep patients on their... [More](#)

Source: <http://www.gnshealthcare.com/>



# Isabel: Diagnostic Decision Support

**enter clinical features**      **synonyms**

age\*

gender  female  male

**Refine search:**  
 travel history:

**show me:**  
 diagnoses  
 causative drugs  
 bioterrorist agents

**Enter clinical features, no negatives, no numbers:**

bilious vomiting	⊗
bloody diarrhea	⊗
fever	⊗
failure to pass meconium	⊗
abdominal distention	⊗
lethargy	⊗
heme-positive stool	⊗
septic shock	⊗

+ add a clinical feature

**get checklist** ➔

[clear search](#)

**diagnoses**      **drugs**

**Results**

sort by: most relevant      action: -select-

show: 10 / all

<input type="checkbox"/> Large Bowel Obstruction			<input type="radio"/> GASTRO
Hirschsprung's Disease			
Meconium Plug Syndrome			

**Why did this diagnosis come up ?**

We matched the terms: bilious vomiting | bilious vomit | green vomit | fever | fevers | feverish | abdominal distention | abdominal distension | lethargy | lethargic | septic shock | failure | failur | pass | meconium

Degree of match between query entered and Isabel database (Not clinical probability): 97%

<input type="checkbox"/> Intussusception			? GASTRO
<input type="checkbox"/> Neonatal Sepsis			? SHOCK
<input type="checkbox"/> Colonic Obstruction			? GASTRO
<input type="checkbox"/> Malrotation / Volvulus			? GASTRO
<input checked="" type="checkbox"/> Intestinal Obstruction			? GASTRO
Necrotizing Enterocolitis			? INFECTION
<input checked="" type="checkbox"/> Cystic Fibrosis			? RESP
CMV Colitis			? GASTRO
Meningococcal Disease			? INFECTION

[view all](#)

feedback:  **submit**

# Errors and Failures in Diagnosis

- 80% of diagnosis errors occur in five categories:
  - Cancer, infections, cardiovascular diseases, GI tract diseases, fractures & other bony diseases
- Primary failure modes:
  - Initial diagnosis (23%)
  - ordering wrong lab/test (19%)
  - incorrect interpretation of test (18%)
  - failing to make a referral (14%)
- Other failure modes occur in 3% of cases:
  - Not communicating lab/test result to practitioner and/or patient
  - Inadequate follow-up plan
  - Incomplete ordering of lab/test

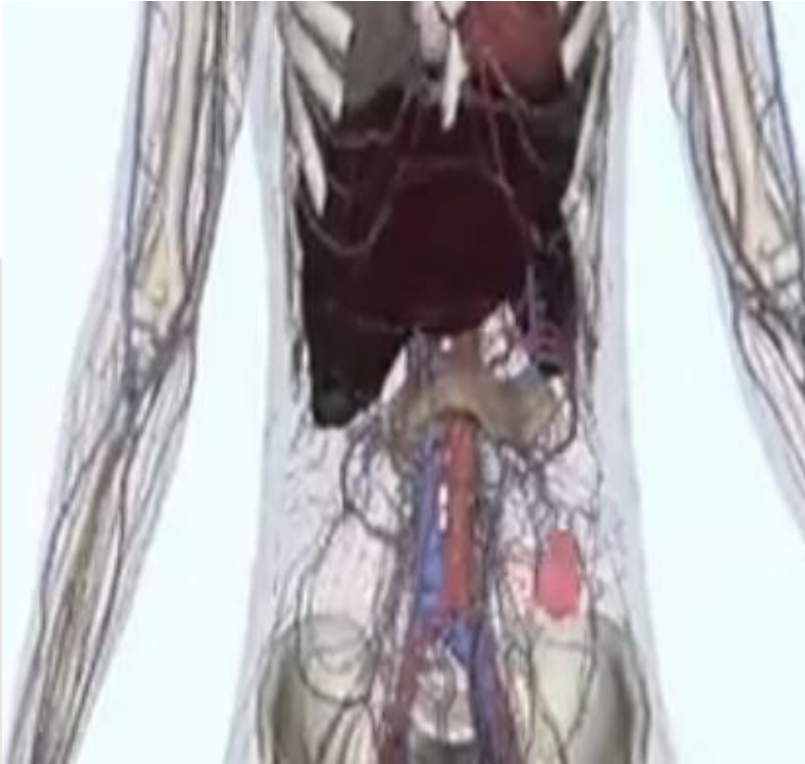
Table 4. Contributory Factors for 5 Process Dimensions<sup>a</sup>

Factor	No. (% of Cases (N = 190))
<b>Patient related (n = 31 [16.3%])</b>	
Failure of patient to provide accurate medical history	14 (7.4)
Lack of clear history from family members in a patient with cognitive dysfunction	8 (4.2)
Patient did not realize that he/she should seek care	6 (3.2)
Failure of communication between practitioner and patient	5 (2.6)
Patient did not realize that he/she should seek care in a more urgent manner	5 (2.6)
<b>Patient-practitioner encounter (n = 150 [78.9%])</b>	
Problems ordering diagnostic tests for further workup	109 (57.4)
Error related to medical history	107 (56.3)
Error related to physician examination performance	90 (47.4)
Failure to review previous documentation	29 (15.3)
<b>Diagnostic tests (n = 26 [13.7%])</b>	
Erroneous clinician interpretation of test and its need for follow-up	9 (4.7)
Considered test result interpretation as nonserious	8 (4.2)
Misinterpretation of clinical test results	7 (3.7)
Being misled by normal history and physical examination findings, laboratory result, or imaging study result	5 (2.6)
Being too focused on one diagnosis or treatment plan	5 (2.6)
No earlier appointment was given	5 (2.6)
Practitioner did not think result was serious enough for admission	5 (2.6)
<b>Follow-up and tracking (n = 28 [14.7%])</b>	
Inadequate test result tracking system	7 (3.7)
No follow-up tracking system	7 (3.7)
Practitioner selected too much time for follow-up	5 (2.6)
Considered condition as nonserious	5 (2.6)
<b>Referrals (n = 37 [19.5%])</b>	
Appropriate expert is not contacted	19 (10)
Considered condition as nonserious	14 (7.4)
Did not believe referral was required	12 (6.3)
Suboptimal weighing of critical piece of history data	10 (5.3)
Lack of knowledge or insufficient practitioner knowledge of relevant condition	5 (2.6)

<sup>a</sup>Each case may have several contributing factors involved.

Source: <http://archinte.jamanetwork.com> – Feb. 25, 2013

# The Intuitive Electronic Medical Record



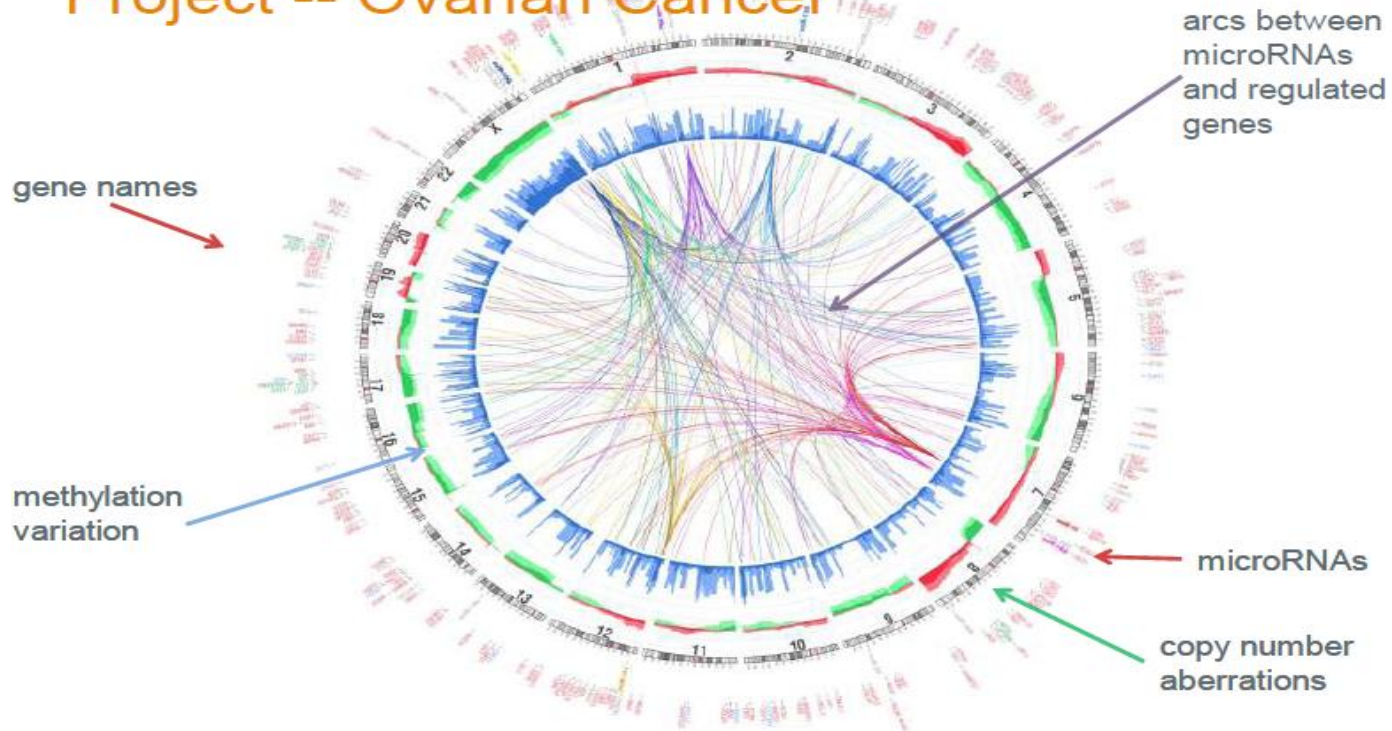
A screenshot of the Epic EMR system interface. The main window displays a 'Lab Report' for a 'MISCELLANEOUS REFERENCE LAB TEST 1 (Order 319241857)'. The report includes a 'Result Information' section with a 'Final result (7/20/2012 12:46 PM)' and a note to see results older than 5 years. Below this is a table for 'MISCELLANEOUS REFERENCE LAB TEST 1' with columns for 'Component' and 'Value'. The table lists 'Reference lab test results' as 'SEE REPT', 'Performing reference laboratory' as 'Disneyland', 'Lab test name' as 'SPICE', and 'Specimen source' as 'BLOOD'. It also shows the 'Resulting Agency' as 'KFH SANTA CLARA - HOMESTEAD LABORATORY' and 'Specimen Collected' on 07/20/12 at 11:51 AM, with 'Last Resulted' on 07/20/12 at 12:46 PM. A 'Result History' section shows the test on 7/20/12. A 'Result Flag Legend' defines codes: AA = Panic / Critical, LL = Low Panic / Review Alert, H = High, L = Low, and A = Abnormal / \*. At the bottom, patient information is shown: 'M Wwww WITSIIGGXCerner', '60 y.o. / Female (5/15/1952)', and 'MRN: 110013926611'. The order details are 'Order MISCELLANEOUS REFERENCE LAB TEST 1 [220319] (Acc#441220209002) (Order 319241857)'. The interface includes a navigation sidebar on the left with options like 'Snapshot', 'Chart Review', and 'History'. The top of the screen shows the Epic logo and various navigation icons. The bottom status bar indicates 'ONE TECHLAB' and the time '9:04 AM'.

Source: IBM and Shutterstock

# Personalized Medicine – Genomics

# The Promise of Genomic Data

## ISB Genomic Data Integration: TCGA Project -- Ovarian Cancer



Source: Institute of Medicine of the National Academies, 2012.

# When Will Genomics Revolutionize Medicine?



- Tomorrow?



- 5 Years?



- 10 Years?

# Complexity of Cancer Genetics



The NEW ENGLAND  
JOURNAL of MEDICINE

FREE PowerPoint® Access  
SPECIAL OFFER »



HOME | ARTICLES & MULTIMEDIA ▾ | ISSUES ▾ | SPECIALTIES & TOPICS ▾ | FOR AUTHORS ▾ | CME ▾

Keyword, Title, Author, or Citation

**ORIGINAL ARTICLE** A Correction Has Been Published >

## Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing

Marco Gerlinger, M.D., Andrew J. Rowan, B.Sc., Stuart Horswell, M.Math., James Larkin, M.D., Ph.D., David Endesfelder, Dip.Math., Eva Gronroos, Ph.D., Pierre Martinez, Ph.D., Nicholas Matthews, B.Sc., Aengus Stewart, M.Sc., Patrick Tarpey, Ph.D., Ignacio Varela, Ph.D., Benjamin Phillimore, B.Sc., Sharmin Begum, M.Sc., Neil Q. McDonald, Ph.D., Adam Butler, B.Sc., David Jones, M.Sc., Keiran Raine, M.Sc., Calli Latimer, B.Sc., Claudio R. Santos, Ph.D., Mahrokh Nohadani, H.N.C., Aron C. Eklund, Ph.D., Bradley Spencer-Dene, Ph.D., Graham Clark, B.Sc., Lisa Pickering, M.D., Ph.D., Gordon Stamp, M.D., Martin Gore, M.D., Ph.D., Zoltan Szallasi, M.D., Julian Downward, Ph.D., P. Andrew Futreal, Ph.D., and Charles Swanton, M.D., Ph.D.  
N Engl J Med 2012; 366:883-892 | March 8, 2012 | DOI: 10.1056/NEJMoa1113205

Comments open through March 14, 2012

Share:

Abstract	<b>Article</b>	References	Citing Articles (250)	Comments (4)	Letters
----------	----------------	------------	-----------------------	--------------	---------

**BACKGROUND**

Intratumor heterogeneity may foster tumor evolution and adaptation and hinder personalized-medicine strategies that depend on results from single tumor-biopsy samples.

[Full Text of Background...](#)

**MEDIA IN THIS ARTICLE**

**FIGURE 1**

**TOOLS**

PDF	E-Mail
Print	Save
Download Citation	Article Alert
Slide Set	Reprints
Supplementary Material	Permissions
	Share/Bookmark

**RELATED ARTICLES**

**EDITORIAL**  
Tumor Heterogeneity and Personalized Medicine  
March 8, 2012 | D.L. Longo

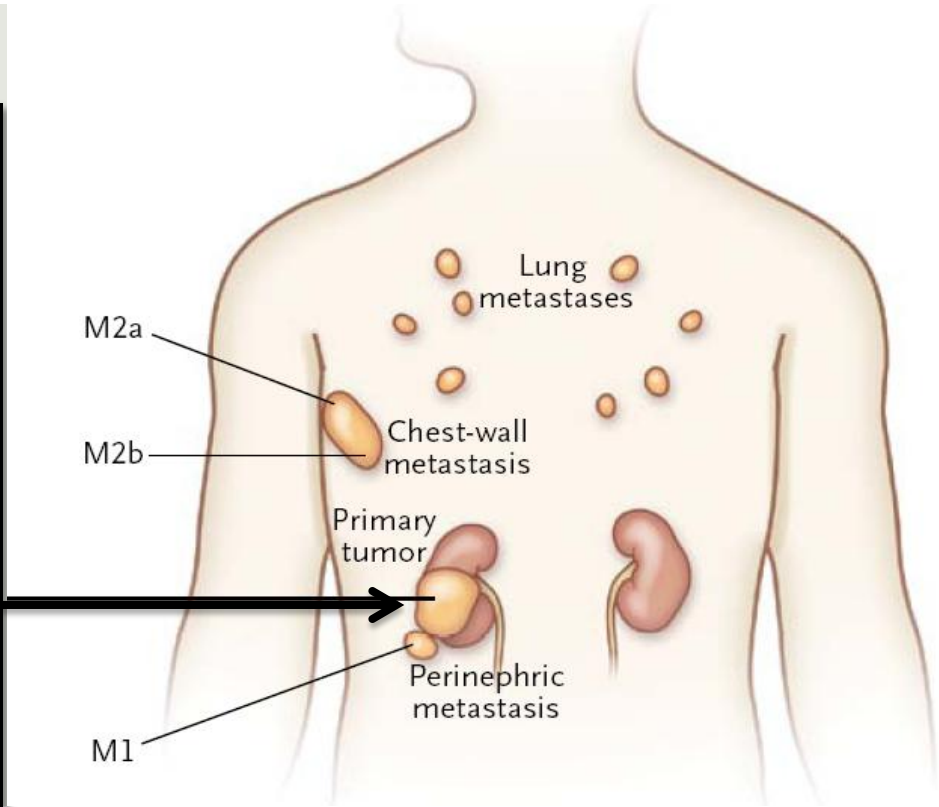
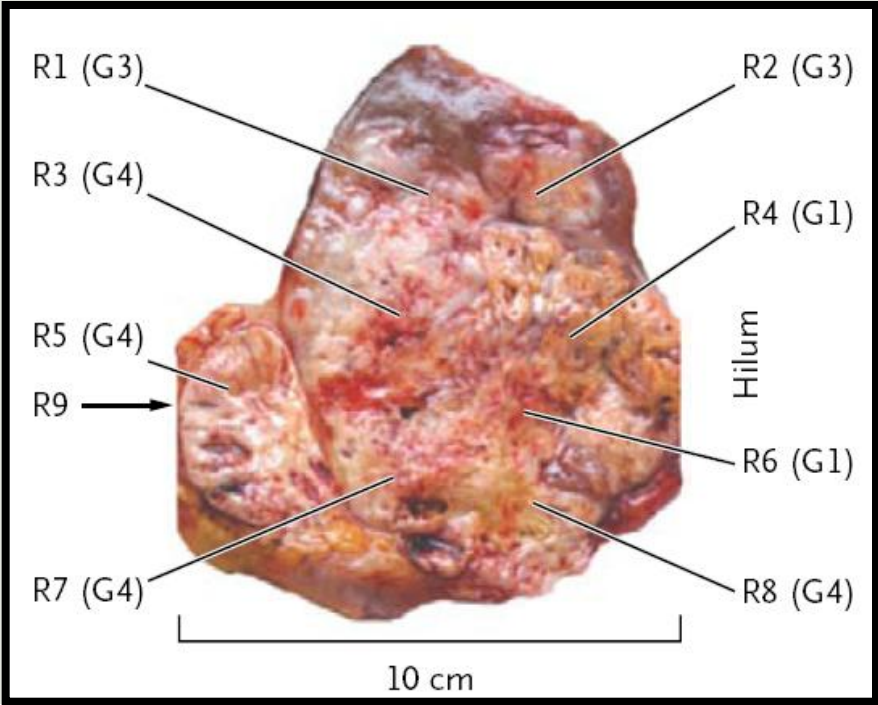
---

**CORRESPONDENCE**  
Intratumor Heterogeneity and Branched Evolution

Source: The New England Journal of Medicine. March 8, 2012 .

# Complexity of Genetic Medicine for Cancer

## Intratumor Heterogeneity



Source: "Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing", New England Journal of Medicine, 2012



# Genetic Contribution to Disease

- The 1,000 Genomes Project cited the genomes of 1,092 individuals from 14 populations.
- The project demonstrated that:
  - Individuals from different populations carry different profiles
  - Low-frequency variants show substantial geographic differentiation

## ARTICLE

doi:10.1038/nature11632

### An integrated map of genetic variation from 1,092 human genomes

The 1000 Genomes Project Consortium\*

By characterizing the geographic and functional spectrum of human genetic variation, the 1000 Genomes Project aims to build a resource to help to understand the genetic contribution to disease. Here we describe the genomes of 1,092 individuals from 14 populations, constructed using a combination of low-coverage whole-genome and exome sequencing. By developing methods to integrate information across several algorithms and diverse data sources, we provide a validated haplotype map of 38 million single nucleotide polymorphisms, 1.4 million short insertions and deletions, and more than 14,000 larger deletions. We show that individuals from different populations carry different profiles of rare and common variants, and that low-frequency variants show substantial geographic differentiation, which is further increased by the action of purifying selection. We show that evolutionary conservation and coding consequence are key determinants of the strength of purifying selection, that rare-variant load varies substantially across biological pathways, and that each individual contains hundreds of rare non-coding variants at conserved sites, such as motif-disrupting changes in transcription-factor-binding sites. This resource, which captures up to 98% of accessible single nucleotide polymorphisms at a frequency of 1% in related populations, enables analysis of common and low-frequency variants in individuals from diverse, including admixed, populations.

Source: "An integrated map of genetic variation from 1,092 human genomes," The 1000 Genomes Project Consortium, NATURE, vol. 491, November 2012.

# Genetic Information Nondiscrimination Act

- Signed into law by President George W. Bush, May 2008
- First major civil rights bill of the century
- Protects against health insurance and employment discrimination based on genetic information



President George W. Bush signs H.R. 493, the Genetic Information Nondiscrimination Act of 2008, Wednesday, May 21, 2008, in the Oval Office.

White House photo by Eric Draper.

Source: <http://www.genome.gov/24519851>

# Privacy Challenges of Genetic Research

## Surnames can be recovered from personal genomes

**Science** AAAS.ORG | FEEDBACK | HELP | LIBRARIANS   **SEARCH** ADVANCED

GUEST | ALERTS | ACCESS RIGHTS | MY ACCOUNT | SIGN IN

**AAAS** NEWS | **SCIENCE JOURNALS** | CAREERS | BLOGS & COMMUNITIES | MULTIMEDIA | COLLECTIONS | **JOIN / SUBSCRIBE**

**Science** The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

Science 18 January 2013:  
Vol. 339 no. 6117 pp. 321-324  
DOI: 10.1126/science.1229566


[< Prev](#) | [Table of Contents](#) | [Next](#)


 [Read Full Text to Comment \(](#)

REPORT

### Identifying Personal Genomes by Surname Inference

Melissa Gymrek<sup>1,2,3,4</sup>, Amy L. McGuire<sup>5</sup>, David Golan<sup>6</sup>, Eran Halperin<sup>7,8,9</sup>, Yaniv Erlich<sup>1,\*</sup>

 [Author Affiliations](#)

 <sup>\*</sup>To whom correspondence should be addressed. E-mail: [yaniv@wi.mit.edu](mailto:yaniv@wi.mit.edu)

#### ABSTRACT

Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state, can be used to triangulate the identity of the target. A key feature of this technique is that it entirely relies on free, publicly accessible Internet resources. We quantitatively analyze the probability of identification for U.S. males. We further demonstrate the feasibility of this technique by tracing back with high probability the identities of multiple participants in public sequencing projects.

Source: Science Magazine, January 18, 2013, <http://www.sciencemag.org/content/339/6117/321.abstract>

The true future of health data lies in its ability to support the safest, highest quality, most individualized care without the constraint of borders or boundaries.

