Mining Clinical Data to build Predictive Model

Peter Szolovits

MIT Computer Science and Artificial Intelligence Lab Professor of Computer Science and Engineering Professor of Health Sciences and Technology

2013 SIAM International Conference on Data Mining Austin, Texas

May 2, 2013







Outline

- The Landscape of Healthcare
 - 1. The Learning Health Care System
 - 2. Universal Electronic Health Records
 - 3. Federal Incentives for Meaningful Use to Improve Quality of Care
 - 4. The Portable ICU Real Time Data Everywhere
 - 5. The \$1,000 Genome
 - 6. "Big Data" Methods for Analysis
- Examples
 - A. Rheumatoid Arthritis
 - B. Intensive Care



2



Inevitable Growth of Healthcare Spending





1. "The Learning Health Care System"

- "one in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and health care" —IOM
- Needs not currently met:
 - Comprehensive collation of all clinical, social, demographic, behavioral, ... data that are now captured in the health care system
 - Routine capture of novel data sources:
 - genomes, gene expression, etc.
 - environmental factors
 - physiological response to life situations
 - (related to fitness and wellness)
 - Technical infrastructure
 - Storage and analysis of truly "big data"
 - Incentives and demonstrations of utility



What is Evidence-Based Medicine?

- Randomized Controlled Clinical Trials
 - E.g., is drug A more effective than drug B for condition X?
 - Narrow selection of patient cases and controls
 - Careful collection of systematically organized data
 - Statistical analysis of outcomes
 - => Statistically significant conclusions
- But:
 - Heterogeneity: Most cases to which RCT results are applied do not fit trial criteria
 - Short Follow-Up: Trials run for limited times, but use is longer
 - Small Samples: Some effects are rare but devastating
- **Instead**: consider every patient's experience as a source of knowledge by which to improve health care



Vioxx and Heart Attacks



"Trends in inpatient stay due to MI were tightly coupled to the rise and fall of prescriptions of COX-2 inhibitors, with an 18.5% increase in inpatient stays for MI when both rofecoxib and celecoxib were on the market (P<0.001). For every million prescriptions of rofecoxib and celecoxib, there was a 0.5% increase in MI (95%CI 0.1 to 0.9) explaining 50.3% of the deviance in yearly variation of MI-related hospitalizations."

Figure 1. Cumulative sum (CUSUM) chart of monthly incidence of hospitalizations due to myocardial infarction from January 1, 1997 to March 30, 2006. Assigning 1997 and 1998 as a baseline period, a target mean of 47.1 hospitalizations due to myocardial infarctions per 100,000 and a standard deviation of 2.8 per 100,000. The threshold for CUSUM was calculated as 1.18, yielding an average-run-length of 50 (dashed red line). An aberration, two standard deviations from baseline, was initially detected in January of 2000, with a CUSUM value of 1.37 and 51.5 myocardial infarction-related hospitalization per 100,000 stays (solid line). doi:10.1371/journal.pone.0000840.g001

Brownstein JS, Sordo M, Kohane IS, Mandl KD (2007) The Tell-Tale Heart: Population-Based Surveillance Reveals an Association of Rofecoxib and Celecoxib with Myocardial Infarction. PLoS ONE 2(9): e840.



2. Universal Electronic Health Records

- Rapid adoption in hospitals, likely driven by 2009 "Stimulus Bill"
 - "Basic EHR systems include ...: patient demographics, patient problem lists, patient medication histories, clinical notes, electronic orders for prescriptions, laboratory results viewing, and imaging results viewing." —ONC
 (Data from http://dashboard.healthit.gov/HITAdoption/?view=0)
 - Positive factors for adoption: large size, urban location and HMO penetration Health information technology adoption in U.S. Acute care hospitals. Zhang NJ, *et al.* J Med Syst. 2013 Apr;37(2):9907





EHR Challenges

- Interoperability
 - Standards vs. evolution
 - Core data to take care of a patient: Demographics, insurance, diagnoses & problem lists, medications, lab values, allergies, recent encounters, images, care plan
 - E.g., HL7-CCD or CCR: text + optional codes (SNOMED, LOINC, ...)
 - Comprehensive data: HL7-RIM
 - Instance of Berners-Lee's Semantic Web issues
- Impediments to health information exchange
- Workflow changes
 - Errors
 - E.g., mortality 2.8% \Rightarrow 6.6% after CPOE system @ Pitt Children's
 - Waste of effort
 - "Cut and Paste" records
- Proprietary lock-in

Han et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics (2005) vol. 116 (6) pp. 1506-12

8



3. Meaningful Use for Quality Improvement

- 2009 Recovery Act defines "Meaningful Use":
 - Use EHR in a meaningful manner (e.g., e-prescribing)
 - Electronic exchange of health information to improve quality
 - Submit clinical quality measures

http://www.healthit.gov/policy-researchers-implementers/meaningful-use

Stage 1 Data Capture & Sharing 2011-12	Stage 2: Advance Clinical Processes 2014	Stage 3: Improved Outcomes 2016
Electronically capturing health information in a standardized format	More rigorous health information exchange (HIE)	Improving quality, safety, and efficiency, leading to improved health outcomes
Using that information to track key clinical conditions	Increased requirements for e- prescribing and incorporating lab results	Decision support for national high- priority conditions
Communicating that information for care coordination processes	Electronic transmission of patient care summaries across multiple settings	Patient access to self-management tools
Initiating the reporting of clinical quality measures and public health information	More patient-controlled data	Access to comprehensive patient data through patient-centered HIE
Using information to engage patients and their families in their care		Improving population health



led.

Reducing Readmissions

- Starting Oct 1, 2012, hospitals are penalized by Medicare if a patient is re-admitted for the same condition within 30 days
 - Initial focus conditions: Acute Myocardial Infarction, Heart Failure, Pneumonia

118 papers on

"predict readmission"

since 2010 in Pubmed

- Risk adjustments for demographics
- Predictive models are worth \$\$\$
 - E.g., Potentially avoidable 30-day he derivation and validation of a predic 2013 Apr 22;173(8):632-8
 - Retrospective study of 10,731 di
 - 2,398 (22.3%) readmitted within 30 days, or a (0.570) judged avoidable
 - Eliminated planned re-admissions and those for other conditions
 - Logistic regression model predicts avoidable re-admission with AUC=0.71, good calibration
 - Important variables: hemoglobin at discharge, oncology service, Na+ level at discharge, procedures performed, non-elective admission, number of admissions in past 12 months, length of stay

10



4. Real Time Data Everywhere

Home health & fitness devices





5. The \$1,000 Genome: Genomic Medicine

- Exponentially falling price of sequencing
 - Archon Genomics X PRIZE \$10M
 - Sequence 100 human genomes in 30 days
 - accuracy, completeness, and haplotype phasing
- The \$1,000,000 interpretation CLARITY Challenge
 - What are the best methods to process the massive amounts of data?
 - Should the results change how a patient's care is managed?
 - How do we present the information so it's understandable and useful?
 - When should we report unexpected, incidental findings?
 - How can we safeguard patient privacy?
 - <u>http://genes.childrenshospital.org</u>
 - Similar techniques, in different combinations



George Church



David Margulies



6. Data Analytics and Big Data

- Volume, Velocity, Variety
- Intel Science and Technology Center on Big Data
 - @ MIT, Brown, Portland State, Stanford, UCSB, U. Tenn, U. Wash
 - Algorithms
 - Sub-linear (sampling), good error estimates
 - Distributed/parallel
 - Incremental machine learning
 - Graphs, sparse matrices
 - Storage and Hardware
 - Move computation to data, not data to computation
 - No caching
 - Visualization
 - Applications
 - · Geographic surveillance, health care, genomics, urban systems, ...



Shift from Knowledge to Data



I2B2: Informatics for Integrating Biology and the Bedside

- Basic driving idea:
 - Phenotype = f(Genotype, Environment)
 - What is *f*?
 - Uses: better diagnostics, personalized medicine/treatment, prevention, research (applied and systems biology)
- Ten-year project at Partners Healthcare
 - Very influential in setting directions for research in CTSA sites
- Principles
 - Use "found data"
 - Patients being treated in the course of the "natural experiments" of disease
 - Clinical records rather than experimental data collection protocols
 - Crimson system for using (de-identified) lab samples acquired during routine care
 - "A million anecdotes are data!" —Zak Kohane
- Focus on predictive modeling.

Sources of "Found Data"

- Codified Data Sets
 - Lab measurements
 - Bedside measurements (vital signs, ...)
 - Prescription orders, pharmacy fulfillment
 - Procedure and billing codes
 - Monitoring data
 - Intensive care
 - Home health
 - Genetics: SNPs, CNVs, Exomes, whole genome sequences
 - Geographic location

- Narrative Data
 - Doctors' and nurses' notes
 - Radiology, pathology, ... reports
 - Discharge summaries
 - Referral letters
 - Blogs, diaries, posts to social media
- Imaging

A. Rheumatoid Arthritis Study

• Liao KP et al., Electronic Medical Records for Discovery Research in Rheumatoid Arthritis, Arthritis Care and Research 62(8) 1120-7, 2010.

- Goal: identify ~1500 patients with 97% specificity who have RA, and collect a blood sample from them to use in GWAS.
 - GWAS is easily corrupted by including inaccurately determined cases
- Simply using billing codes yields PPVs of 57% to 66% in best reported studies

Example of an i2b2 Web Client Query

Typical Machine Learning Approach

- Generate a large variety of features
 - Billing codes
 - Measured lab values
 - Medications and dosages
 - Frequency of doctors' visits and hospitalizations
 - Total "fact load"
 - NLP on notes and discharge summaries to find other evidence of the above
 - e.g., results and prescriptions elsewhere are not in codified data, but are often mentioned in narrative reports
 - NLP for judgmental facts, e.g., "joint erosions" on x-ray
 - Mathematical transforms of the above; e.g., log for skewed distributions
- Feature selection
- Machine learning algorithm
 - In our case, Penalized Logistic Regression

Table 3. Variables selected for the complete algorithm
(narrative and codified EMR data) from the logistic
regression in order of predictive value*

regression in order of prodictive value		
Variable	Standardized regression coefficient	Standard error
Positive predictors		
NLP RA	1.11	0.48
NLP seropositive	0.74	0.26
ICD-9 RA normalized†	0.71	0.23
ICD-9 RA	0.66	0.44
NLP erosions	0.46	0.29
Codified RF negative	0.36	0.36
NLP methotrexate	0.3	0.34
Codified anti-TNF [‡]	0.29	0.3
NLP anti-CCP positive	0.27	0.25
NLP anti-TNF§	0.2	0.36
NLP other DMARDs	0.13	0.34
Negative predictors		
ICD-9 JRA	-0.98	0.9
ICD-9 SLE	-0.57	1.09
NLP PsA	-0.51	0.74
 * EMR = electronic medical recessing; RA = rheumatoid arthrocation of Diseases, Ninth Revise TNF = anti-tumor necrosis citrullinated peptide; DMARDs drugs; JRA = juvenile rheumate erythematosus; PsA = psoriation + ICD-9 RA normalized = ln (new week apart). ‡ Codified anti-TNF = etanerce not available in our EMR). § NLP anti-TNF = adalimumabe 	ecord; NLP = natural ritis; ICD-9 = Internat sion; RF = rheumatoi factor; anti-CCP = disease-modifying oid arthritis; SLE = s arthritis. o. of ICD-9 RA codes p pt and infliximab (ada o, etanercept, and infli	language pro- tional Classifi- d factor; anti- = anti-cyclic antirheumatic ystemic lupus per subject ≥1 alimumab was iximab.

Table 4. Comparison of performance characteristics from validation of the complete classification algorithm (narrative and codified) with algorithms containing codified-only and narrative-only data*				
Model	RA by algorithm or criteria, no.	PPV (95% CI), %	Sensitivity (95% CI), %	Difference in PPV (95% CI), %†
Algorithms				
Narrative and codified (complete)	3,585	94 (91–96)	63 (51-75)	Reference
Codified only	3,046	88 (84-92)	51 (42-60)	6 (2–9)‡
NLP only	3,341	89 (86–93)	56 (46-66)	5 (1-8)‡
Published administrative codified criteria				
≥3 ICD-9 RA codes	7,960	56 (47-64)	80 (72-88)	38 (29-47)‡
\geq 1 ICD-9 RA codes plus \geq 1 DMARD	7,799	45 (37–53)	66 (57-76)	49 (40–57)‡

* The complete classification algorithm was also compared with criteria for RA used in published administrative database studies. RA = rheumatoid arthritis; PPV = positive predictive value; 95% CI = 95% confidence interval; NLP = natural language processing; ICD-9 = International Classification of Diseases, Ninth Revision; DMARD = disease-modifying antirheumatic drug.

+ Difference in PPV = PPV of complete algorithm - comparison algorithm or criteria.

‡ Significant difference in PPV compared with the complete algorithm.

Obtaining Genomic and Custom Lab Data: Merging and Anonymizing Data Sets

- · Lab analysis of discarded samples
- Re-identification risk mitigated by IRB oversight and data use agreement
- · Crimson acts as "honest broker"

Genetic Study Based on this Patient Selection

Kurreeman et al. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. American Journal of Human Genetics (2011) vol. 88 (1) pp. 57-69

- Population:
 - 1,515 cases of RA
 - 4,575 potential cases selected at 95% specificity
 - 1,480 controls
- Rheumatoid Arthritis
 - ~60% of disease variability is inherited
 - 30 known gene loci explain ~20% of variation
 - Most patients studies have been European, + for ACPA and/or RF
- Our goal: study other ethnic groups, and ACPA- patients
- Procedure:
 - Broad genotyped 192 ancestry-informative markers, 29 SNPs from 27 RA risk loci
 - Measured ACPA +/- status

Asterisks indicate TNFAIP3 SNP rs6920220 and CCL21 SNP rs951005. EU indicates individuals of European descent from our EHR cohort. GWAS represents samples from the previously published GWAS meta-analysis.²²

PTPN22

2.7

Genetic Study Results

• EHR and genetically derived ancestry concordance:

Europeans	98%
Africans	94%
East Asians	78%
Other (Hispanics)	52%

- Odds ratios for individual SNPs in general agreement with meta-analysis of much larger, experimentally collected cohort
- Defined a Genetic Risk Score (GRS) to aggregate the small effects of individual SNPs $GRS = \sum_{i=1}^{n} w_i X_i \text{ where } X_i \in \{0, 1, 2\}$
- Excellent agreement with distributions from large meta-analysis
- Similar differences between cases and controls in all four populations (significant in all but East Asians, which had smallest cohorts)

Figure 3. Distribution of the Aggregate Genetic-Risk Score from 29 RA Risk Alleles in ACPA+ Cases versus Controls Samples represented in the respective panels are (A) controls and ACPA+ cases in our EHR study and controls (orange lines) and seropositive (ACPA+ and or RF+) and healthy individuals from the GWAS meta-analysis, all of European descent (EU) (blue lines); (B) controls and ACPA+ cases of African descent (AF) (green lines), (C) controls and ACPA+ cases of East Asian descent (AS) (purple lines), and (D) controls and ACPA+ cases of Hispanic descent (HIS) (gray lines).

1 L

The Idea is Becoming Common ...

BMJ Open 2012;2:e000850 doi:10.1136/bmjopen-2012-000850

Pharmacology and therapeutics

Hypnotics' association with mortality or cancer: a matched cohort study

@ ⊕ ⊕ OPEN ACCESS

Daniel F Kripke¹, Robert D Langer², Lawrence E Kline¹

Design Longitudinal electronic medical records were extracted for a one-to-two matched cohort survival analysis.

For groups prescribed 0.4–18, 18–132 and >132 doses/year, HRs (95% CIs) were 3.60 (2.92 to 4.44), 4.43 (3.67 to 5.36) and 5.32 (4.50 to 6.30), respectively, demonstrating a dose–response association. HRs were elevated in separate analyses for several common hypnotics, including zolpidem, temazepam, eszopiclone, zaleplon, other benzodiazepines, barbiturates and sedative antihistamines. Hypnotic use in the upper third was associated with a significant elevation of incident cancer; HR=1.35 (95% CI 1.18 to 1.55). Results were robust within groups suffering each comorbidity, indicating that the death and cancer hazards associated with hypnotic drugs were not attributable to pre-existing disease.

- MIMIC-II data set contains comprehensive data on 32,535 patients, with over 40,000 ICU stays
 - Saeed et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit Care Med* (2011) vol. 39 (5) pp. 952-960
 - Codes: Demographics, which ICU, co-morbidities, labs, meds, I/O, procedures, ICD9 & DRG codes, microbiology, orders
 - Narratives: Nursing notes, doctors' notes, specialist reports, discharge summaries, …
 - Vitals: heart rate, blood pressure, temperature
 - **Monitors**: 5-minute summaries of monitored values
 - Waveforms: high-resolution waveforms for up to 8 channels of data (on ~10% of stays) "almost Big Data"

Using MIMIC Data to Build Predictive Models

http://dspace.mit.edu/handle/1721.1/46690

- Mortality
 - Comparison to SAPS II
 - Daily Acuity Scores
 - Real-time Acuity Scores (real-time risk assessment)
- Other clinical events
 - pressor weaning
 - intra-aortic balloon pump weaning
 - onset of septic shock
 - acute kidney injury
- Data set (MIMIC 2, earlier snapshot)
 - 10,066 patients: 7,048 development, 3,018 validation
 - selected cases with adequate data
 - excluded neurological and trauma cases

Cleaning the data—half the research time

- Missing values
 - Some values are not measured for some clinical situations
 - Failures in data capture process
- Episodically measured variables
 - Extrapolate
- (ouenbeu Unclear/undefined clinical states
- Imprecise timing of meds, ...
- Partially measured i/o
- Proxies: e.g., which ICU \Rightarrow what disease
- Select subset of data with enough data points!

What Kinds of Models to Build?

- Patient state depends on pathophysiology
 - Genetic complement, environmental exposures, pathogens, auto-regulatory mechanisms, treatments, ...
- Possible formalism:
 - POMDP's, but intractable
 - Graphical models (Bayes Nets, Influence Diagrams, etc.), but require many independence assumptions
 - Simple models: Cox proportional hazard, naïve Bayes, linear/logistic regression
- Derived variables can summarize essential contributions of dynamic variation
 - integrals, slopes, ranges, frequencies, etc.
 - Transformed variables: inverse, abs, square, square root, log-abs, abs deviation from mean, log abs deviation, ...

Summary of Mortality Models

Figure 5-27: AUC versus day, patients with ICU stays ≥ 5 days (validation data)

Models for Therapeutic Opportunities and Risks

Prediction	AUC
Weaning from vasopressors within next 12 hours, remain off for 4 hours	0.809
Pressor weaning + Survival	0.825
Weaning from Intra-Aortic Balloon Pump	0.816
Onset of Septic Shock	0.843
Acute kidney injury	0.742

- Predictions are not as accurate as mortality models, but still impressive
- Using acuity score instead of such specific models is worse
 - E.g., Vasopressor weaning 0.679 vs. 0.809

But where to go from here?

- Include data from narrative records: discharge summaries, radiology/ pathology/... reports, nursing/doctor notes, ...
 - Requires natural language processing
- Bottom-up clustering of pathophysiological states
 - Provides a more abstract description of the patient; reduces "curse of dimensionality"
 - Inspiration: Mitchell Cohen *et al.*, "Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis", Crit Care 2010
 - By time, organ system, major therapy
 - e.g., Kshetri MEng thesis
- Relation to knowledge-based ("expert systems") models

Clustering Snapshots <u>http://groups.csail.mit.edu/medg/ftp/kshetri/Kshetri_MEng.pdf</u>

 ~10,000 patients x ~12,000 possible features/patient ≈ 1M rows (sparse), 30 min snapshots ⇒ 11 clusters

Finding better Abstractions

Find better ways to represent and reason with massive medical data

Prediction using multi-layer abstraction

Slide from Rohit Joshi

Standardize All Variables

Measures deviations from normal ranges

Standardized x

40

Abstraction via Radial Domain Folding (1)

• For each focus k, compute magnitude of deviation of each relevant variable x_i for patient j

- For each focus k, cluster the M_{kj} , order clusters by aggregate magnitude (6 clusters)
- Jaccard code the d_{ij} and cluster them (sparse, 8 clusters)
- We now have an abstract patient description along each focus $< m_k$, $d_k > (48 \text{ possibilities})$

Abstraction via Radial Domain Folding (2)

- Perform a 2nd level aggregation
 - Patient severity from focus "severities" (cluster order)
 - Direction abstraction derived from focus direction clusters
- Cluster patient severities and the higher-level directions
- Details in Joshi and Szolovits. Prognostic physiology: modeling patient severity in intensive care units using radial domain folding. AMIA Annual Symposium proceedings (2012) vol. 2012 pp. 1276-83
 - Won the 2012 AMIA student paper competition

Foci and Features

- Selected ~10,000 patients in MIMIC II database (excluded pediatrics, trauma)
- Over 12 domain foci; many clinical variables per focus
- Over 1 million chart events (after binning into 1 hour windows)

Focus	Features in each Focus	
Kidney	Creatinine, BUN, BUN2Cr, UrineOut/Hr/Kg,	
Liver	AST, Alt, TBili, Dbili, Albumin, tProtein	
Cardio	MAP, HR, CVP, BPSys, BPDias, Cardiac Index,	
Respiration	RR, SpO2, FiO2Set, PEEPSet, TidVolSet, SaO2, PIP, M	linVent,
Hematology	Hematocrit, Hgb, Platelets, INR, WBC, RBC, PT	
Electrolytes	Na, Mg, K, Ca, Glucose,	
Acid-base	Art CO2, Art PaCO2, Art pH, Art BE	
General	GCS, Age, temp	
Medication type	Diuretic, Antiarrhythmic, Antiplatelet, Sympathomi	metic,
Chronic	AIDS, HematMalig, Metacarcinoma	
Electrocardio (EKG)	PVC, Rhythm types, Ectopic frequency,	Slide from Rohit Josh

Layer 1: Focus-severity Graph using RDF

Layer 2: Disease Severity using RDF

cluster mortality rate

Learned clusters are closely related to mortality !!

Slide from Rohit Joshi

Mortality Prediction on the Test Data

Our method outperforms customized SAPS-II score Slide from Rohit Joshi

Modeling Patient-state transition

Slide from Rohit Joshi

In(organ_transition) ~ current_state(organ) + current_state(otherorgans) + Iatest_transition _trend(otherorgans) + duration_in_current_state(organ) + Slide from Rohit Joshi duration_since_last_change(otherorgans)

Forecasting the transitions of organ systems

Cardiovascular

Respiratory

Friday, May 3, 13

Recap

- Changing State of Healthcare
 - 1. The Learning Health Care System
 - 2. Universal Electronic Health Records
 - 3. Federal Incentives for Meaningful Use to Improve Quality of Care
 - 4. The Portable ICU Real Time Data Everywhere
 - 5. The \$1,000 Genome
 - 6. "Big Data" Methods for Analysis
- Opportunities for Modeling
 - Personalized Medicine based on Prediction
 - Basis for Scientific Analysis
 - Machine Learning Challenges
 - Big Data Challenges

Thanks

- Funding from
 - NIH-National Library of Medicine (i2b2)
 - NIH-National Institute of Biomedical Imaging and Bioengineering (MIMIC)
 - ONC SHARP 4 project (Secondary Use of Clinical Data)
- RA collaborators:
 - Tianxi Cai, Susanne Churchill, Vivian Gainer, Sergey Goryachev, Elizabeth Karlson, Isaac Kohane, Fina Kurreema, Katherine Liao, Shawn Murphy, Robert Plenge, Soumya Raychaudhuri, Qing Zeng-Treitler, etc.
- NLP collaborators
 - Özlem Uzuner, Bill Long, Anna Rumshisky, Guergana Savova, Marzyeh Ghassemi, Yuan Luo, Andreea Bodnari, Tawanda Sibanda
- Modeling collaborators
 - Rohit Joshi, Bill Long, Caleb Hug, Kanak Kshetri
- My research group
 - <u>http://medg.csail.mit.edu</u>

The END

Supplementary Information

Phase 1 Meaningful Use

- 2009 Recovery Act defines "Meaningful Use":
 - Use EHR in a meaningful manner (e.g., e-prescribing)
 - Electronic exchange of health information to improve quality
 - Submit clinical quality measures
- Phase 1 Requirements for Hospitals: (from http://www.cms.gov/EHRIncentivePrograms/)

Computerized provider order entry (CPOE)	Record and chart changes in vital signs
Drug-drug and drug-allergy interaction checks	Record smoking status for patients 13 years or older
Record demographics	Report hospital clinical quality measures to CMS or States
Implement one clinical decision support rule	Provide patients with an electronic copy of their health information, upon request
Maintain up-to-date problem list of current and active diagnoses	Provide patients with an electronic copy of their discharge instructions at time of discharge, upon request
Maintain active medication list	Capability to exchange key clinical information among providers of care and patient- authorized entities electronically
Maintain active medication allergy list	Protect electronic health information