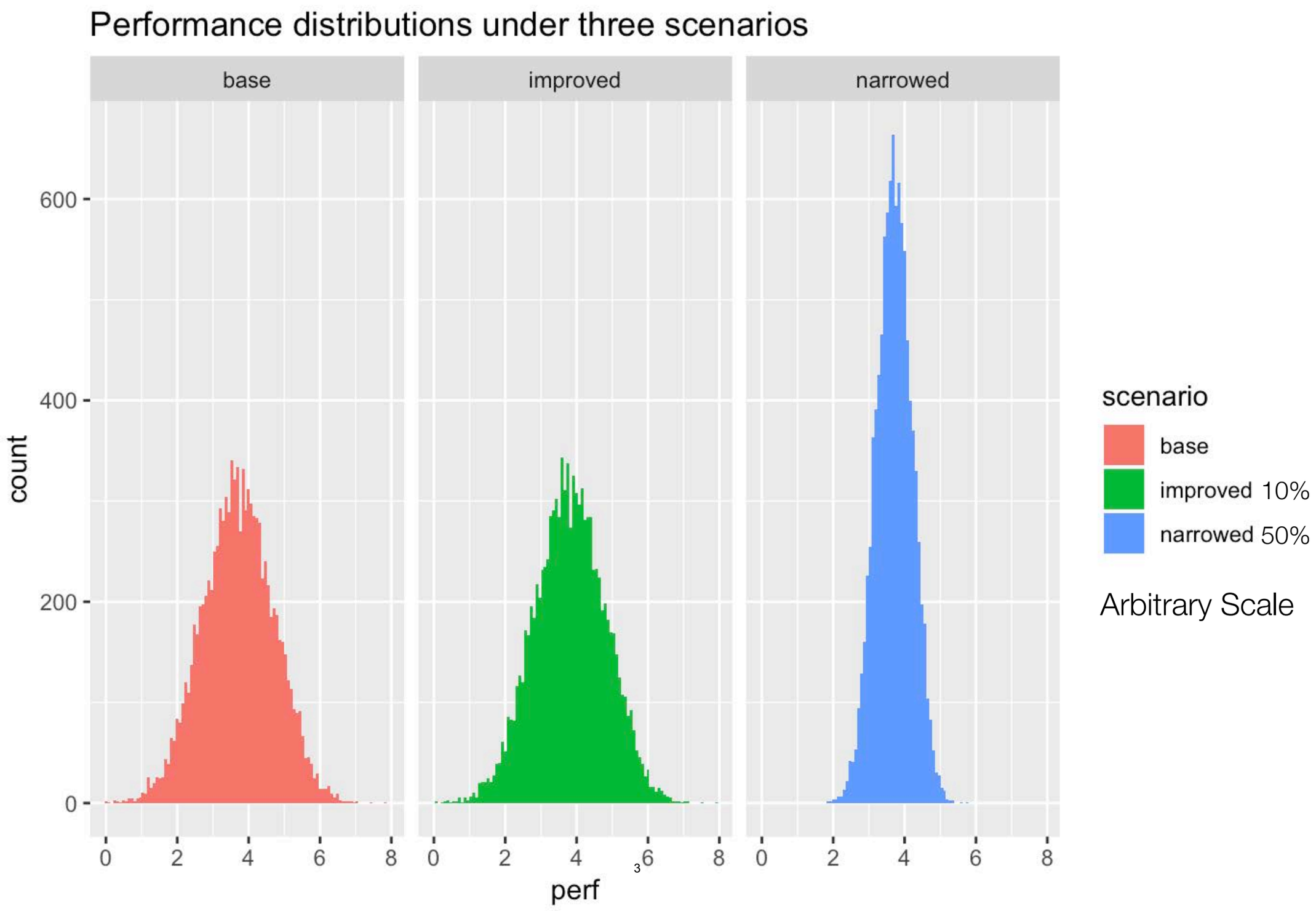
Workflow

Massachusetts Institute of Technology

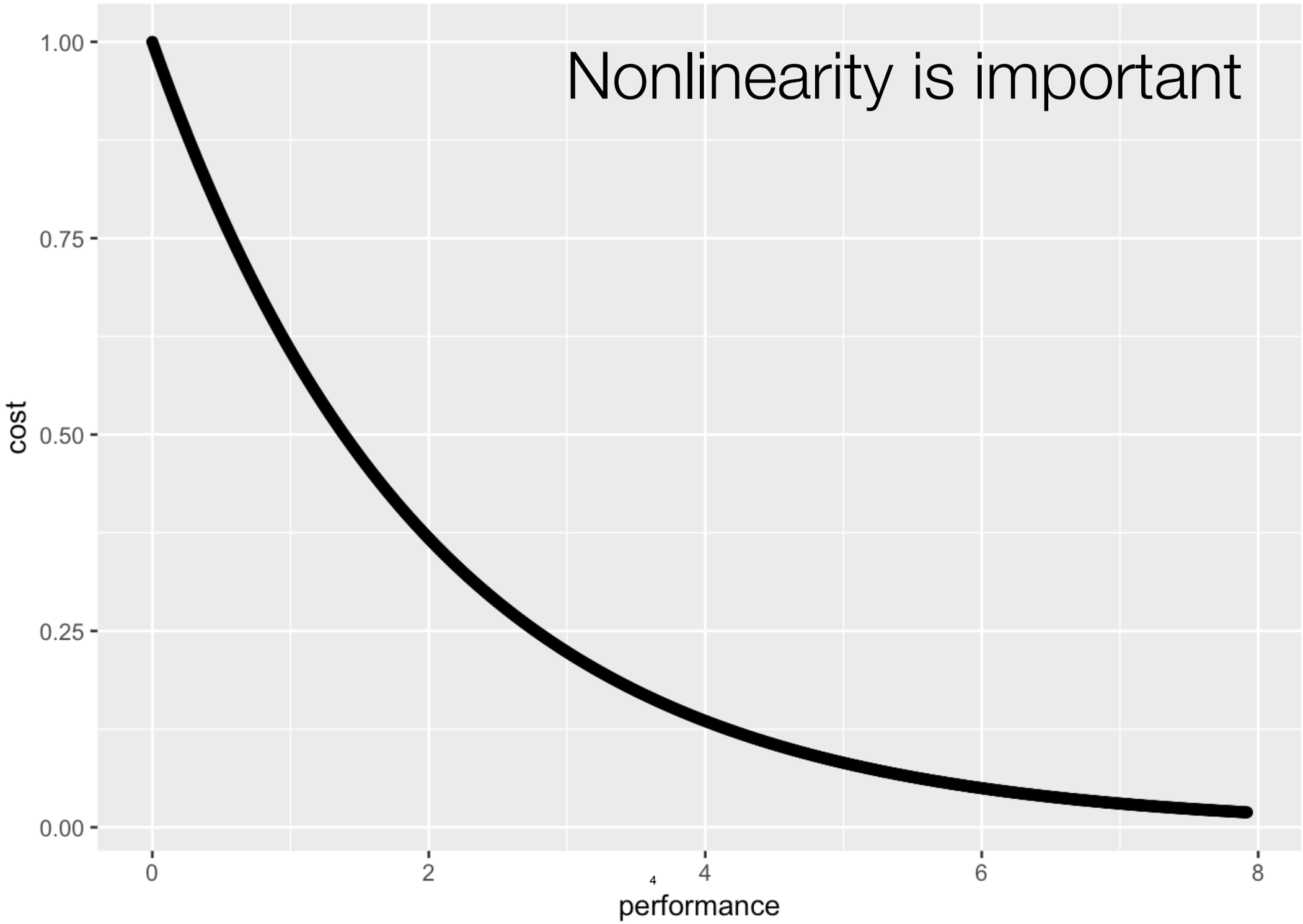
How to Improve Medical Care, Overall

- "Expert Systems" idea: understand what world-class experts do, and provide decision support to raise others' performance to that level
  - *improves average*
- "Protocol" idea: get everyone to treat similar patients in similar ways
  - *reduces variance*
- Which is better?
  - Depends on "loss function"
  - If worst performance is disproportionately more costly than best performance is less costly, then it's more important to eliminate the worst

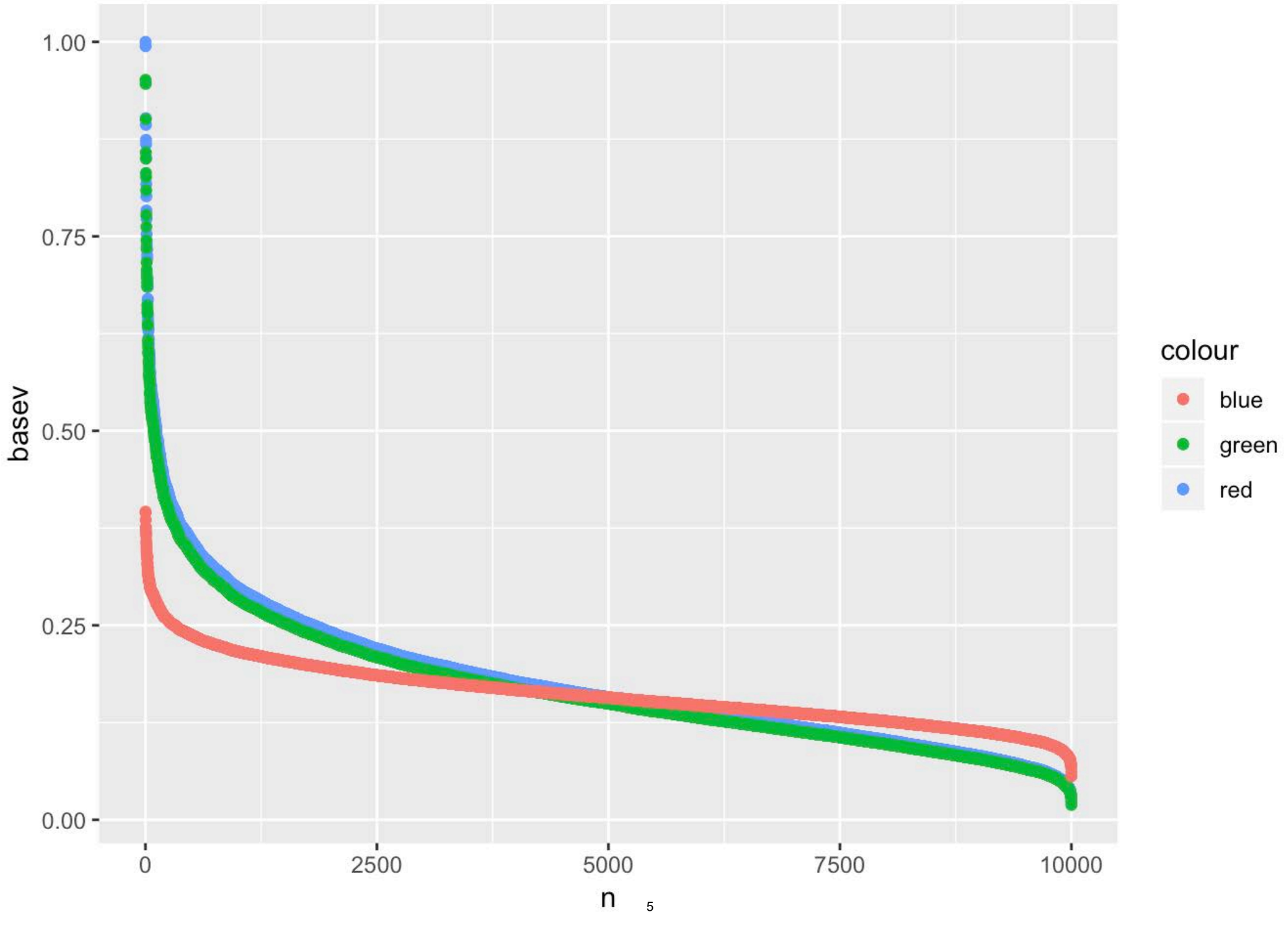
Hypothetical Clinician Performance



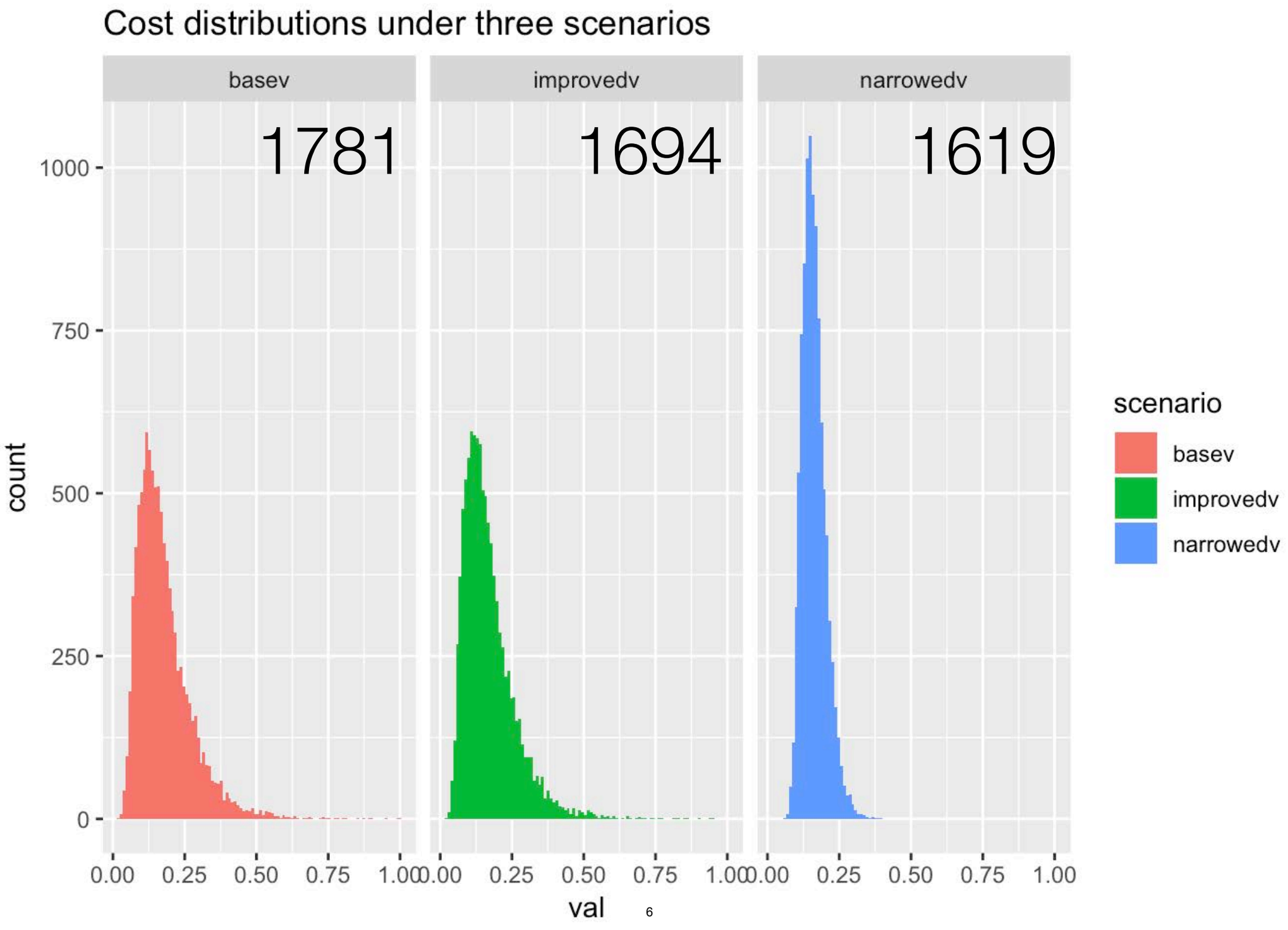
Hypothetical Cost Function



Cost of *n*-th Action Under Three Scenarios



Hypothetical Costs Under Three Scenarios



How to Narrow the Performance Distribution?

- Guidelines and Protocols
  - Learned bodies prescribe appropriate methods to diagnose and treat patients
  - Often based on meta-analysis of clinical trials results
    - Usual caveats about lack of appropriate trials for most conditions

2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol

A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines

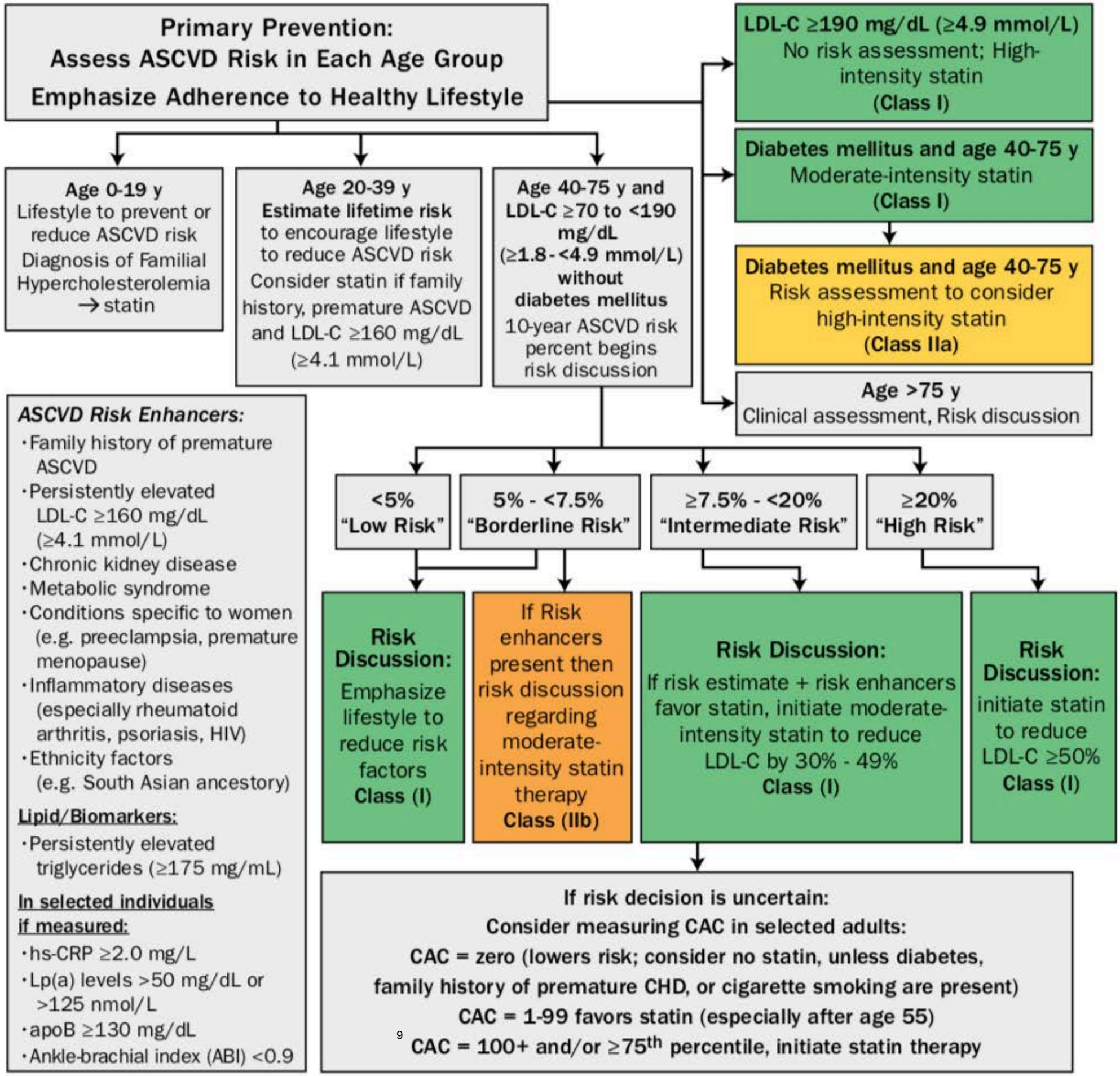
Nov 2018

“Take-Home Messages to Reduce Risk of Atherosclerotic Cardiovascular Disease (ASCVD) through Cholesterol Management

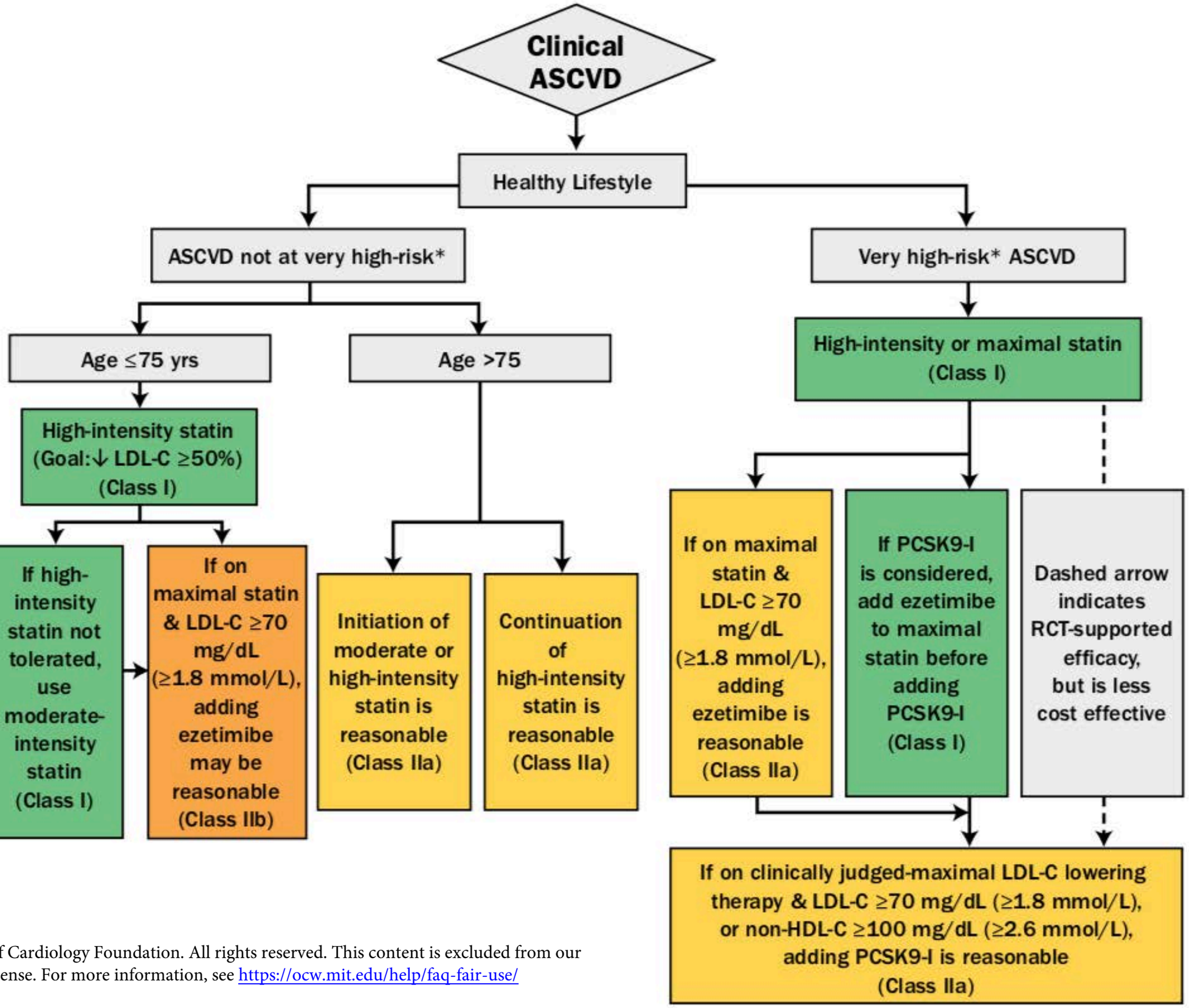
1. In all individuals, emphasize heart-healthy lifestyle across the life-course
2. In patients with clinical ASCVD, reduce low-density lipoprotein cholesterol (LDL-C) with high-intensity statin therapy or maximally tolerated statin therapy
3. In very high-risk ASCVD, use a LDL-C threshold of 70 mg/dL (1.8 mmol/L) to consider addition of nonstatins to statin therapy
4. In patients with severe primary hypercholesterolemia (LDL-C level ≥190 mg/dL [≥4.9 mmol/L]), without calculating 10-year ASCVD risk, begin high-intensity statin therapy without calculating 10-year ASCVD risk
5. In patients 40 to 75 years of age with diabetes mellitus and LDL-C ≥70 mg/dL (≥1.8 mmol/L), start moderate-intensity statin therapy without calculating 10-year ASCVD risk
6. In adults 40 to 75 years of age evaluated for primary ASCVD prevention, have a clinician–patient risk discussion before starting statin therapy
7. In adults 40 to 75 years of age without diabetes mellitus and with LDL-C levels ≥70 mg/dL (≥1.8 mmol/L), at a 10-year ASCVD risk of ≥7.5%, start a moderate-intensity statin if a discussion of treatment options favors statin therapy
8. In adults 40 to 75 years of age without diabetes mellitus and 10-year risk of 7.5% to 19.9% (intermediate risk), risk-enhanc- ing factors favor initiation of statin therapy (see #7)
9. In adults 40 to 75 years of age without diabetes mellitus and with LDL-C levels ≥70 mg/dL- 189 mg/dL (≥1.8-4.9 mmol/L), at a 10-year ASCVD risk of ≥7.5% to 19.9%, if a decision about statin therapy is uncertain, consider measuring CAC
10. Assess adherence and percentage response to LDL-C–lowering medications and lifestyle changes with repeat lipid measurement 4 to 12 weeks after statin initiation or dose adjustment, repeat- ed every 3 to 12 months as needed

Primary Prevention

People without clinical disease



Secondary Prevention in Patients with Clinical ASCVD



*Very high-risk includes a history of multiple major ASCVD events or 1 major ASCVD event and multiple high-risk conditions (**Table 4** on following page).

Very High-Risk for Future ASCVD Events

Very High Risk includes a history of multiple major ASCVD events or one major ASCVD event and multiple high-risk conditions.

Major ASCVD Events
Recent acute coronary syndrome (within the past 12 months)
History of myocardial infarction (other than recent acute coronary syndrome event listed above)
History of ischemic stroke
Symptomatic peripheral arterial disease (history of claudication with ankle brachial index <0.85, or previous revascularization or amputation)
High-Risk Conditions
Age ≥65 years
Heterozygous familial hypercholesterolemia
History of prior coronary artery bypass surgery or PCI outside of the major ASCVD event(s)
Diabetes Mellitus
Hypertension
Chronic kidney disease (eGFR 15-59 mL/min/1.73 m2)
Current smoking
Persistently elevated LDL-C (LDL-C ≥100 mg/dL (≥2.6 mmol/L)) despite maximally tolerated statin therapy and ezetimibe
History of congestive heart failure

Where to Find Guidelines

- AHRQ's National Guideline Clearinghouse
  - Since 1997, but shut down by current administration in July 2018
- Guideline Central (https://www.guidelinecentral.com), ~2K guidelines
  - Assessment of Therapeutic Effectiveness
  - Counseling
  - Diagnosis
  - Evaluation
  - Management
  - Prevention
  - Rehabilitation
  - Risk Assessment
  - Screening
  - Technology Assessment
  - Treatment

Example Guidelines from GuidelineCentral

Assessment and Therapeutic Effectiveness	Calculators
Risk reduction of prostate cancer with drugs or nutritional supplements	4Ts Score for Heparin-Induced Thrombocytopenia
Stem cell transplantation in multiple myeloma	A-a O2 Gradient (need for massive transfusion in trauma)
Stem cell transplantation in myelodysplastic syndromes and acute myeloid leukemia	ABCD2 Score for TIA (risk of stroke after a TIA)
Stem cell transplantation in primary systemic amyloidosis	ACR-EULAR Gout Classification Criteria
The role of liver resection in colorectal cancer metastases	ADAPT Protocol for Cardiac Event (2-hours risk of cardiac event for chest pain)
Optimal chemotherapy for recurrent ovarian cancer	APACHE II Score (ICU mortality)
Radionuclide therapy for neuroendocrine malignancies	APGAR Score (neonates 1 and 5 minutes after birth)

https://www.guidelinecentral.com/summaries/#link=https://www.guidelinecentral.com/summaries/categories/assessment-of-therapeutic-effectiveness/&activeTab=#summary-view-category

https://www.guidelinecentral.com/calculators/

Top-Down vs. Bottom-Up

- Guidelines
  - Typically developed by “learned societies”, usually MDs
  - Choice based on clinical importance, controversy, “pet” ideas, …
- Care Plans
  - Individualized to specific patient
  - Developed by nurse taking care of that patient
- Clinical Pathways
  - Generalization of Care Plans
  - Typically developed by hospitals, combining multidisciplinary sources
    - Guidelines, Nursing experience, Clinical Trials, …
  - Choice based on need to standardize care locally, sometimes in response to errors

Typical Care Plans

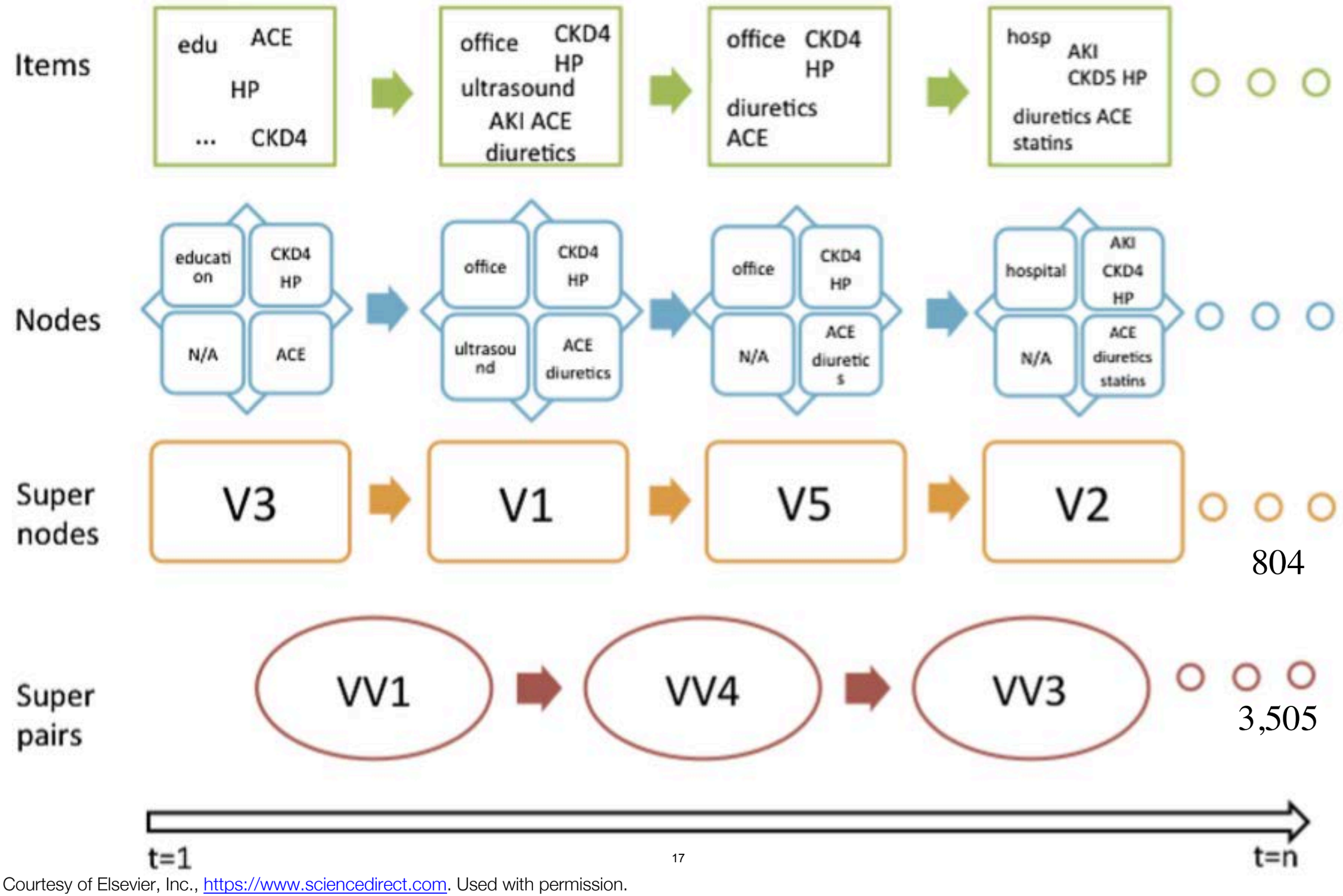
Care Plans
Activities Care Plan
Admission Care Plan
Adult Failure to Thrive Care Plan
Alcohol Withdrawal Care Plan
Allergic Rhinitis Care Plan
Altered Cardiac Output Care Plan
Amputation Care Plan
Anasarca Care Plan
Anemia Care Plan
Angina Care Plan
Anticoagulant Care Plan
Aphasia Care Plan
Arthritis Care Plan
Asthma Management Plan for School Nurse
Behavior Problem Care Plan
Benign Prostate Hypertrophy Care Plan
Breast Feeding Careplan
Cancer Care Plan
Cardiomegaly Care Plan
Cellulitis
Cerebral Palsy Care Plan

https://www.careplans.com/pages/lib/default.aspx?cid=6

Mining Clinical Pathways: Representation

- An *event* is a visit, with a purpose and sets of:
  - procedures,
  - medications: {Angiotensin converting enzyme (ACE) inhibitors, Angiotensin receptor blockers (ARB), diuretics, and statins
  - diagnoses: {CKD stage 1 to stage 5, AKI, hypertension, diabetes, end stage renal disease (ESRD)}
- These events are abstracted into *supernodes*
  - each captures a unique combination of events associated with some visit
- Each patient then has a *visit sequence*, a time-ordered list of supernodes describing successive visits
- To support a two-step Markov analysis, aggregate visits into *super pairs* of two successive supernodes.

Visit History as a Markov Chain



Mining Clinical Pathways: Clustering

- Computer max of the length of common subsequences between each pair of visit sequences
- dist(x, y) = |x| + |y| - 2 LCS(x, y)
- hierarchic clustering into distinct subgroups (31, in their case)

How Useful is This?

- Many subgroups, with 10–158 samples
- Limited data about each visit
  - e.g., no labs, few diagnoses and medication classes
- Complex transition graphs need human interpretation
- Models what *is* done, not what *should be* done
- (but this is a common problem)

Decision support from local data: Creating adaptive order menus from past clinician behavior

Jeffrey G.Klann, Peter Szolovits, Stephen M.Downs, GuntherSchadow

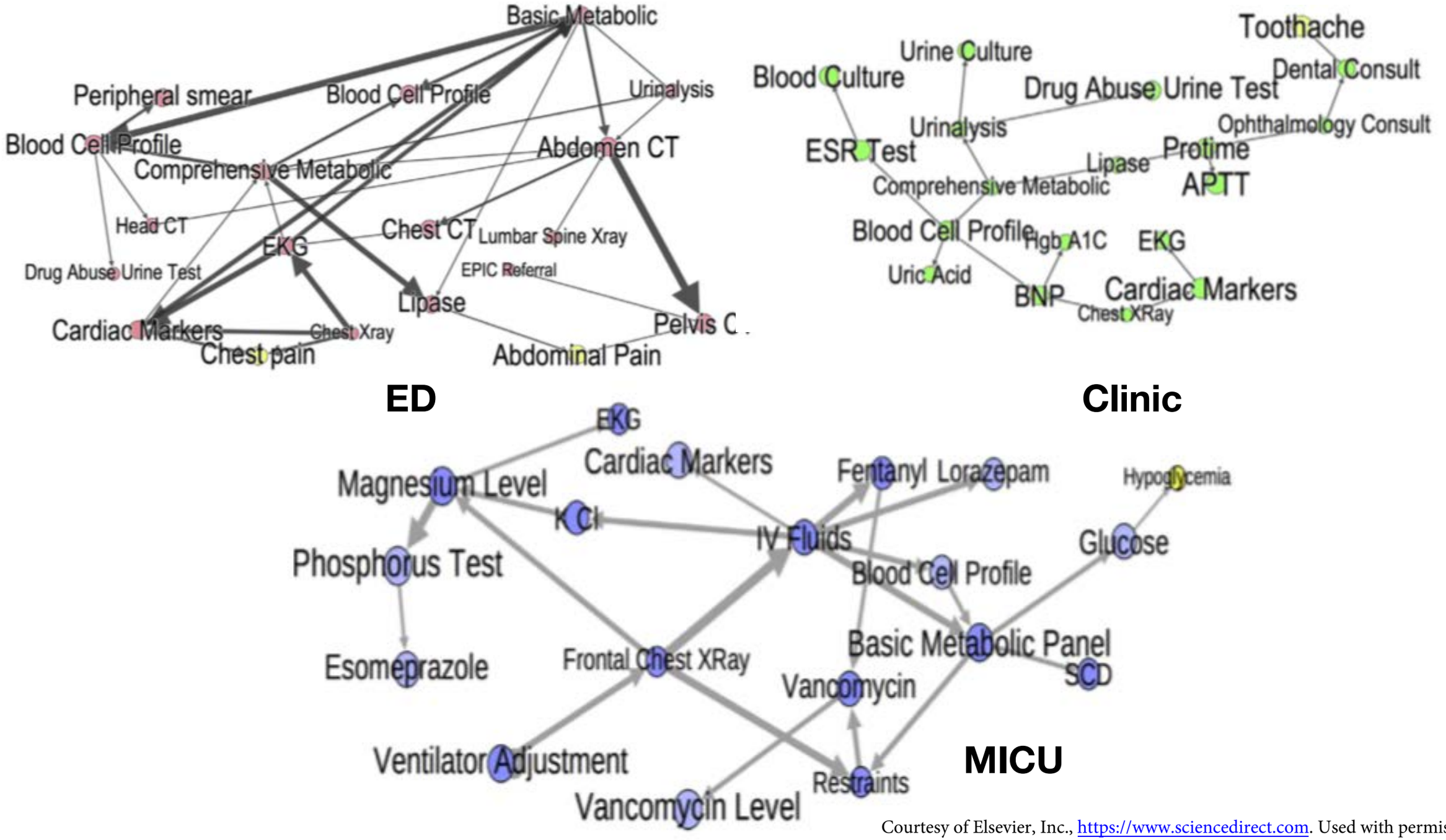
- Clinical Issues
  - back pain in the emergency department (n=9,228)
  - inpatient pregnancy (n=4,843)
  - hypertension in the Urgent Visit Clinic (n=1821)
  - altered mental state in the intensive care unit (n=1,546)
- 3 years of encounters from Regenstrief Clinic
- Data for each domain:
  - 40 most frequent orders (low granularity; e.g., drug, but not dose, for medications)
  - 10 most frequent co-occurring diagnoses

Modeling Clinician Behavior for Decision Support

“”

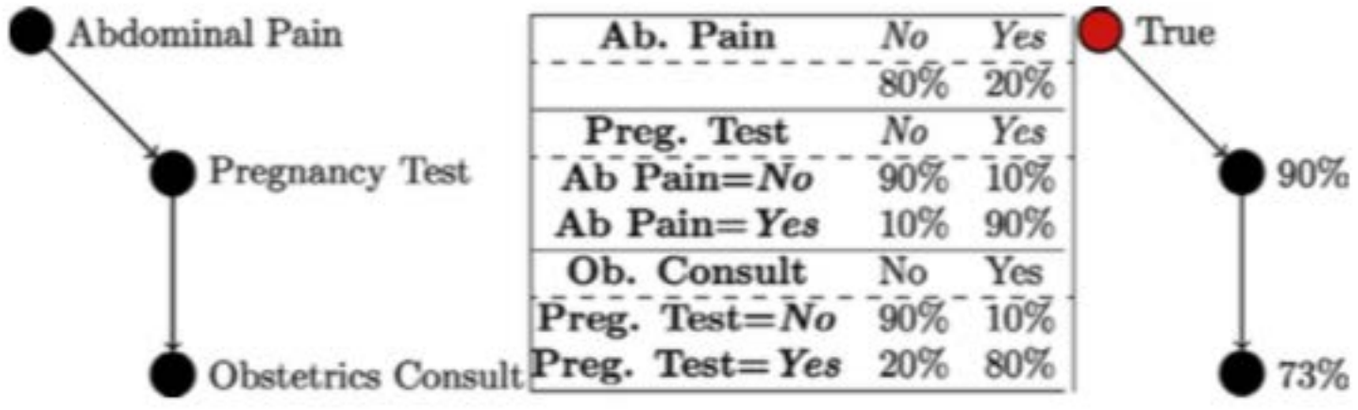
- Wisdom of the Crowd
  - average behavior of many physicians is usually much better than any individual physician
- Like Amazon’s recommendation system: “people who bought this camera also bought this case”
  - too little context
  - inattention to transitive associations
- Automate learning of decision support rules
- Deal with more complex cases than what expert panels typically cover
- Bayesian Network model
  - Diagnoses
  - Possible orders
  - Evidence (from orders already completed)
- Tetrad’s “Greedy Equivalence Search” algorithm to build BN

MICU, Clinic, and ED Networks



MICUNode/label size is proportional to AUC, and edge weight is an approximation of the strength of the relationship. Here, notice the logical clusters and intuitively correct relationships.

Iterative Treatment Suggestion



An example Bayesian Network (left), the Conditional Probability Tables associated with it (middle), and the posterior probabilities given the evidence of ‘Abdominal Pain’ (right).

- Update BN probabilities of possible orders that have not been done
- Present them in descending probability order to clinicians
- Iterate until user ends session

ITS Evaluation by Simulation from Models

Actual context of diagnoses, orders placed; use models to predict next orders

- AUC of action included in recommendations
- Position on recommendation list
- Compare to Association Rule Mining

Pregnancy, Inpatient

	Name	AUC	#
BEST	Sitz Bath	1.00	1.0
	Cold Pack	1.00	1.1
	Naloxone Inj	1.00	1.2
	Lung Exercise	0.99	1.1
	Morphine (PCA)	0.99	2.0
	Ext. UC Monitor	0.99	1.0
	Ibuprofen	0.98	1.1
	Ext. FHT Monitor	0.97	1.1
	Docusate Na	0.96	1.2
	I&O Monitoring	0.94	1.2
WORST	NPO	0.73	1.5
	IV Lock	0.73	9.8
	Syphilis Screen	0.73	9.5
	Ice Chips	0.72	15.8
	IV Fluids	0.71	1.1
	Drugs Urine Test	0.71	27.8
	Oxytocin Protocol	0.68	23.8
	Type and Screen	0.65	13.2
	Lortab 5/500	0.60	2.9
	Morphine	0.50	22.7

Wright A, *et al.* J Am Med Inform Assoc 2016;23:1068–1076. doi:10.1093/jamia/ocw005, Research and Applications

Analysis of clinical decision support system malfunctions: a case series and survey

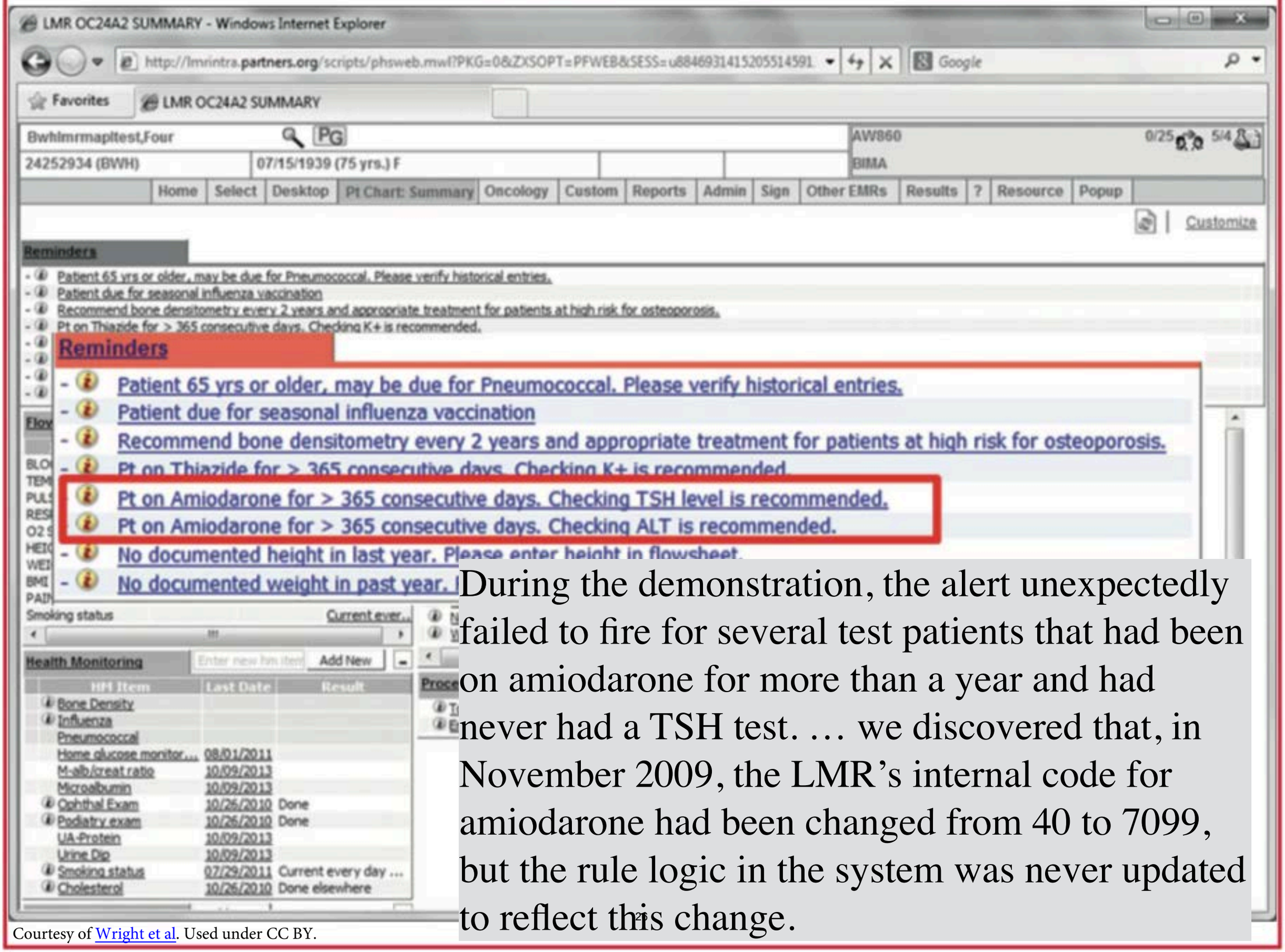
RECEIVED 6 November 2015
REVISED 7 January 2016
ACCEPTED 12 January 2016
PUBLISHED ONLINE FIRST 28 March 2016

Adam Wright,[1,2,3,*] Thu-Trang T Hickman,[1] Dustin McEvoy,[3] Skye Aaron,[1] Angela Ai,[1] Jan Marie Andersen,[1] Salman Hussain,[1,4] Rachel Ramoni,[2,5] Julie Fiskio,[1] Dean F Sittig,[6] and David W Bates[1,2,3]

ABSTRACT

Objective To illustrate ways in which clinical decision support systems (CDSSs) malfunction and identify patterns of such malfunctions.
Materials and Methods We identified and investigated several CDSS malfunctions at Brigham and Women's Hospital and present them as a case series. We also conducted a preliminary survey of Chief Medical Information Officers to assess the frequency of such malfunctions.
Results We identified four CDSS malfunctions at Brigham and Women's Hospital: (1) an alert for monitoring thyroid function in patients receiving amiodarone stopped working when an internal identifier for amiodarone was changed in another system; (2) an alert for lead screening for children stopped working when the rule was inadvertently edited; (3) a software upgrade of the electronic health record software caused numerous spurious alerts to fire; and (4) a malfunction in an external drug classification system caused an alert to inappropriately suggest antiplatelet drugs, such as aspirin, for patients already taking one. We found that 93% of the Chief Medical Information Officers who responded to our survey had experienced at least one CDSS malfunction, and two-thirds experienced malfunctions at least annually.
Discussion CDSS malfunctions are widespread and often persist for long periods. The failure of alerts to fire is particularly difficult to detect. A range of causes, including changes in codes and fields, software upgrades, inadvertent disabling or editing of rules, and malfunctions of external systems commonly contribute to CDSS malfunctions, and current approaches for preventing and detecting such malfunctions are inadequate.
Conclusion CDSS malfunctions occur commonly and often go undetected. Better methods are needed to prevent and detect these malfunctions.

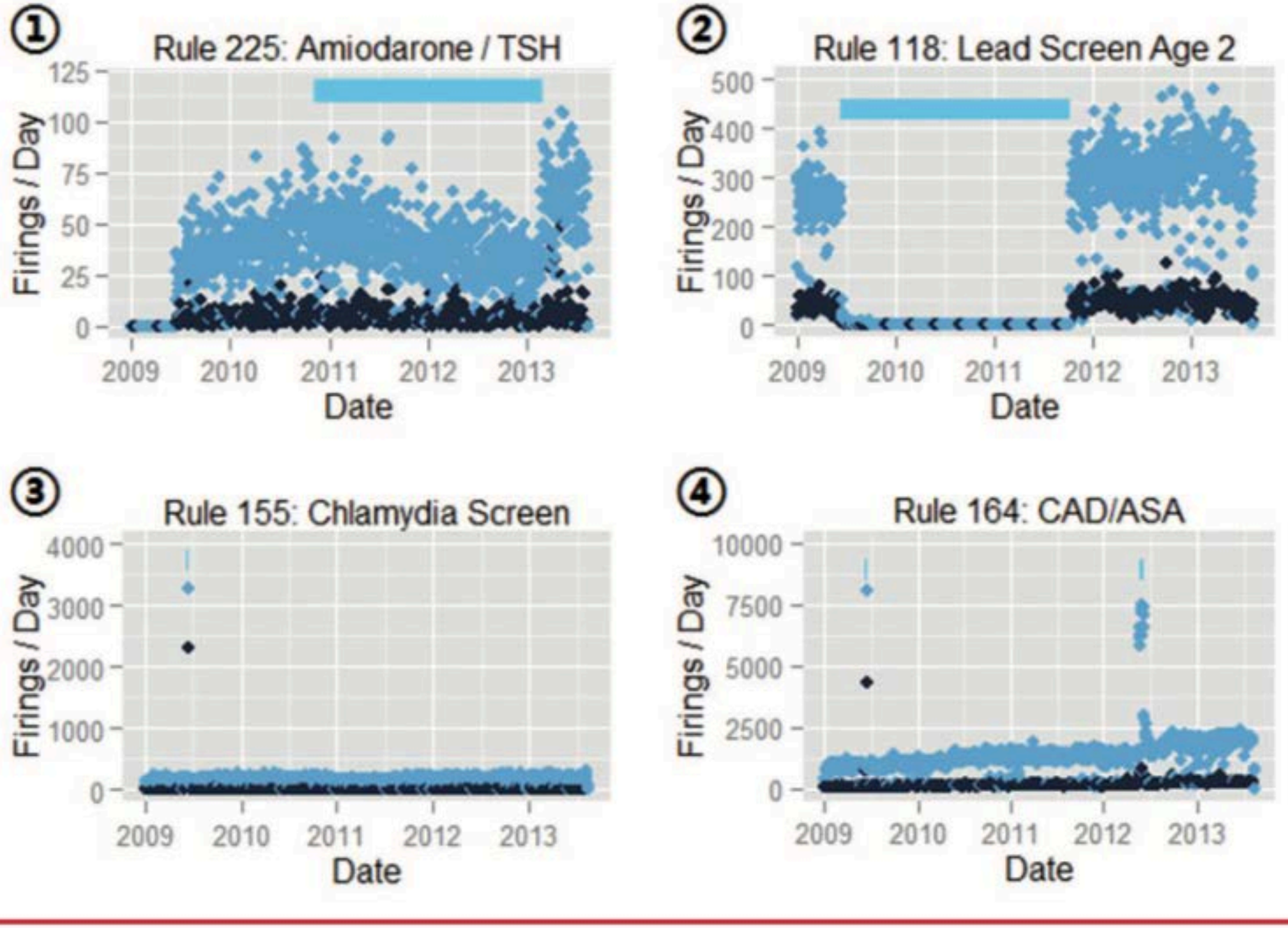
Figure 1: Laboratory monitoring reminders for amiodarone in the Partners Healthcare longitudinal medical record (LMR). The main screen of the LMR is shown in the background, with the reminders enlarged and the amiodarone reminders highlighted in a box.



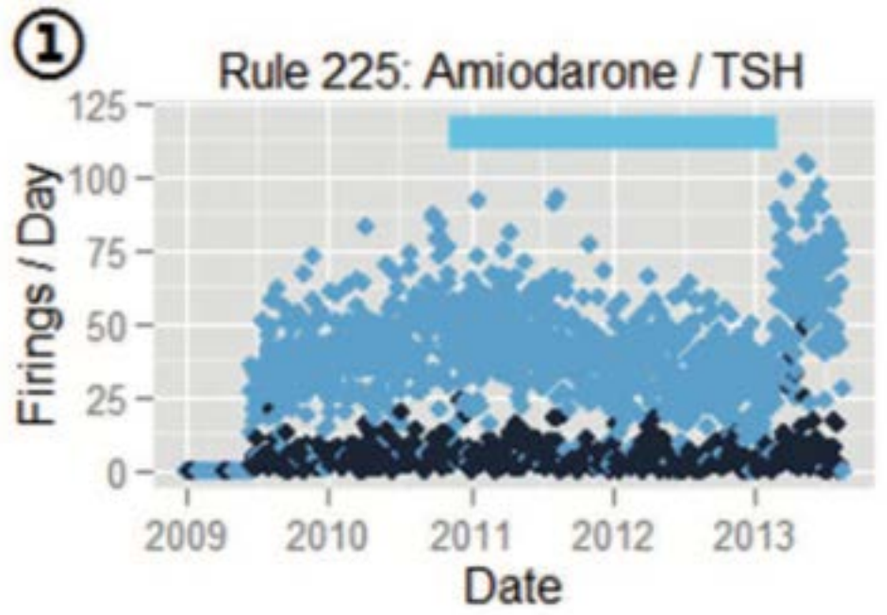
During the demonstration, the alert unexpectedly failed to fire for several test patients that had been on amiodarone for more than a year and had never had a TSH test. … we discovered that, in November 2009, the LMR's internal code for amiodarone had been changed from 40 to 7099, but the rule logic in the system was never updated to reflect this change.

201 Existing Alerts

Figure 3: Firing rate of four alerts at Brigham and Women's Hospital over a 5-year period (weekend days are represented by darker dots, and weekdays are represented by lighter dots), with anomalies indicated (superimposed horizontal bars show anomalous periods).

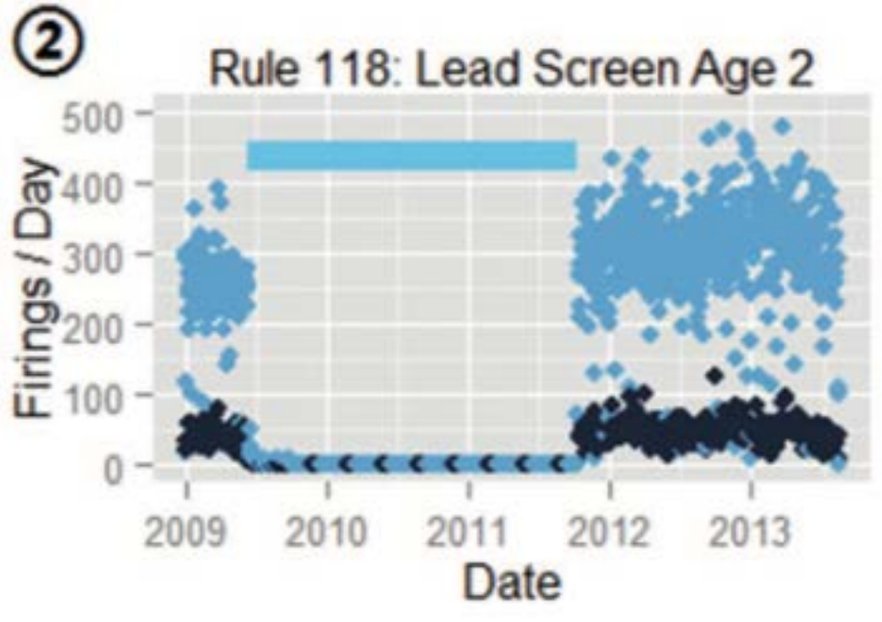


Amioderone



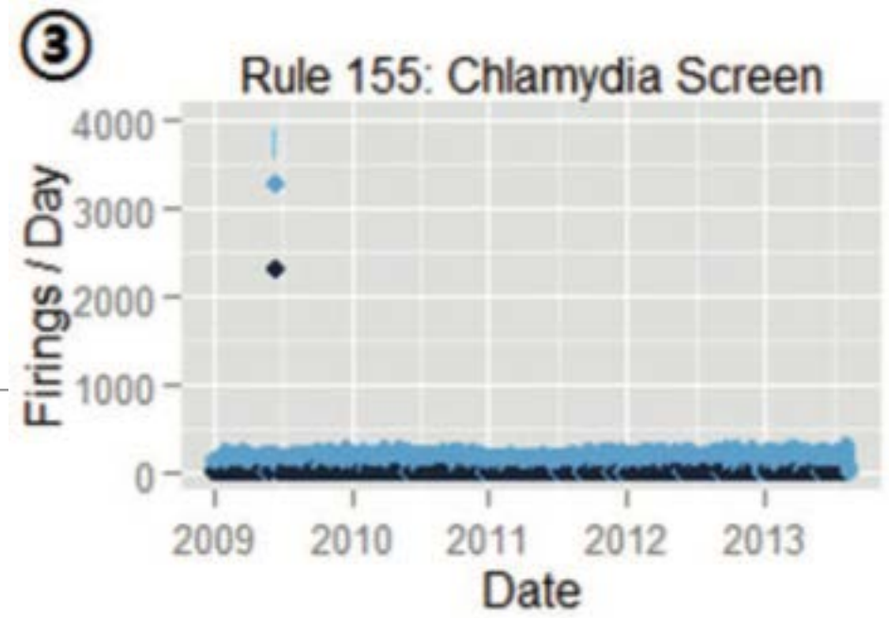
- Because the alert does not fire until a patient has been on amiodarone for at least a year, there was no observable effect for the first year, and then the rate of alerting subtly fell as some patients were taken off amiodarone (with the old code 40) and others were started on amiodarone (with the new internal LMR code 7099). The abrupt increase in the alert firing rate for the amiodarone/TSH test alert at the end of the blue bar in Figure 3 represents when the alert logic was corrected

Lead Screening



- No similar discontinuity for screening 1, 3, and 4-year-olds
- “The audit log suggested that several changes to the lead screening test alert rule were made around the times when the alert stopped firing and then restarted; however, because of a software issue in the audit logging routine, it was not possible to reconstruct the sequence of rule changes or the specific dates when individual changes occurred.”
- Apparently, inadvertent addition of two incomplete clauses to the rule (gender and smoking status) caused it never to fire.
- “176 708 lead screening test alerts were not generated during the 850-day period”

Chlamydia Screen



- Code “clean-up” led to accidental over-firing of an irrelevant rule
- “… record of a healthy 2-month-old boy that contains numerous duplicate reminders, including suggestions that the physician order mammograms, Pap smears, pneumococcal vaccination, and cholesterol screening, and suggestions that the patient be started on several medications, all of which should not apply to this young, healthy, male patient.
- “the alert fired 5950 times during the period that the malfunction occurred compared with the 332 times it was expected to fire”

- Can we automate such monitoring?

Change-Point Detection to Monitor Rule Firings

- Dynamic Linear Model with Seasonality

The DLM models a sequence of real-valued observations $\{y_t: t = 1, 2, \cdots\}$ using a sequence of real-valued hidden state vectors $\{x_t: t = 1, 2, \cdots\}$ of dimension d. The dynamics of the model is captured by:

$$y_t = Fx_t + v, \quad v \sim N(0, V), \quad x_t = Gx_{t-1} + w, \quad w \sim N(0, W). \quad (1)$$

where G is a transition matrix that models the change in the hidden state over time, and F is an emission matrix that reflects the expression of observations y_t given the current x_t. Both transition and emission are stochastic and corrupted by a zero-mean Gaussian noise (w and v) with covariance W and V. At the beginning ($t = 0$), we assume the hidden state $x_0 \sim N(m_0, C_0)$, where m_0 and C_0 is the mean and covariance matrix of x_0 respectively.

Liu, S., Wright, A., Sittig, D. F., & Hauskrecht, M. (2017). Change-Point Detection for Monitoring Clinical Decision Support Systems with a Multi-Process Dynamic Linear Model. (Vol. 2017, pp. 569–572). Presented at the Proceedings IEEE International Conference on Bioinformatics and Biomedicine, IEEE. http://doi.org/10.1109/BIBM.2017.8217712

Seasonality

- Decompose x_t into multiple parts:
  - a baseline (u_t) defining the mean
  - a slope (l_t) defining the trend of the mean
  - a seasonal component (s_t) defining the change in the mean for each phase (a day in a week) of a seasonal cycle (a week); p = length of cycle
  - [t]p = (t + p – 1) mod p + 1 that maps the time to its corresponding phase

$$x_t = \left(u_t, l_t, s^{([t]_p)}, s^{([t-1]_p)}, \ldots, s^{([t-p+2]_p)}\right)^T.$$

Multi-Process Dynamic Linear Model

- Multiple DLMs represent different various normal and abnormal behaviors
- Let $M_t^{(i)}$ be a random variable indicating whether model i is driving the time series at time t and generating y_t, and M_t be a vector composed of $M_t^{(i)}$ for all i.
- $Y_t = \{y_u: u = 1, 2, ..., t\}$ is the time series of observations up to t
- Probability that i drives the time series before observation y_t is $p(M_t^{(i)} = 1 | Y_{t-1})$, and after is $p(M_t^{(i)} = 1 | Y_t)$. This can help detect change
- Three basic models
  - MS (stable)
  - MAO (additive outlier)
  - MLS (level shift)
- $p(M_t^{(MLS)} = 1 | Y_{t+1})$ is considered the *change point score*

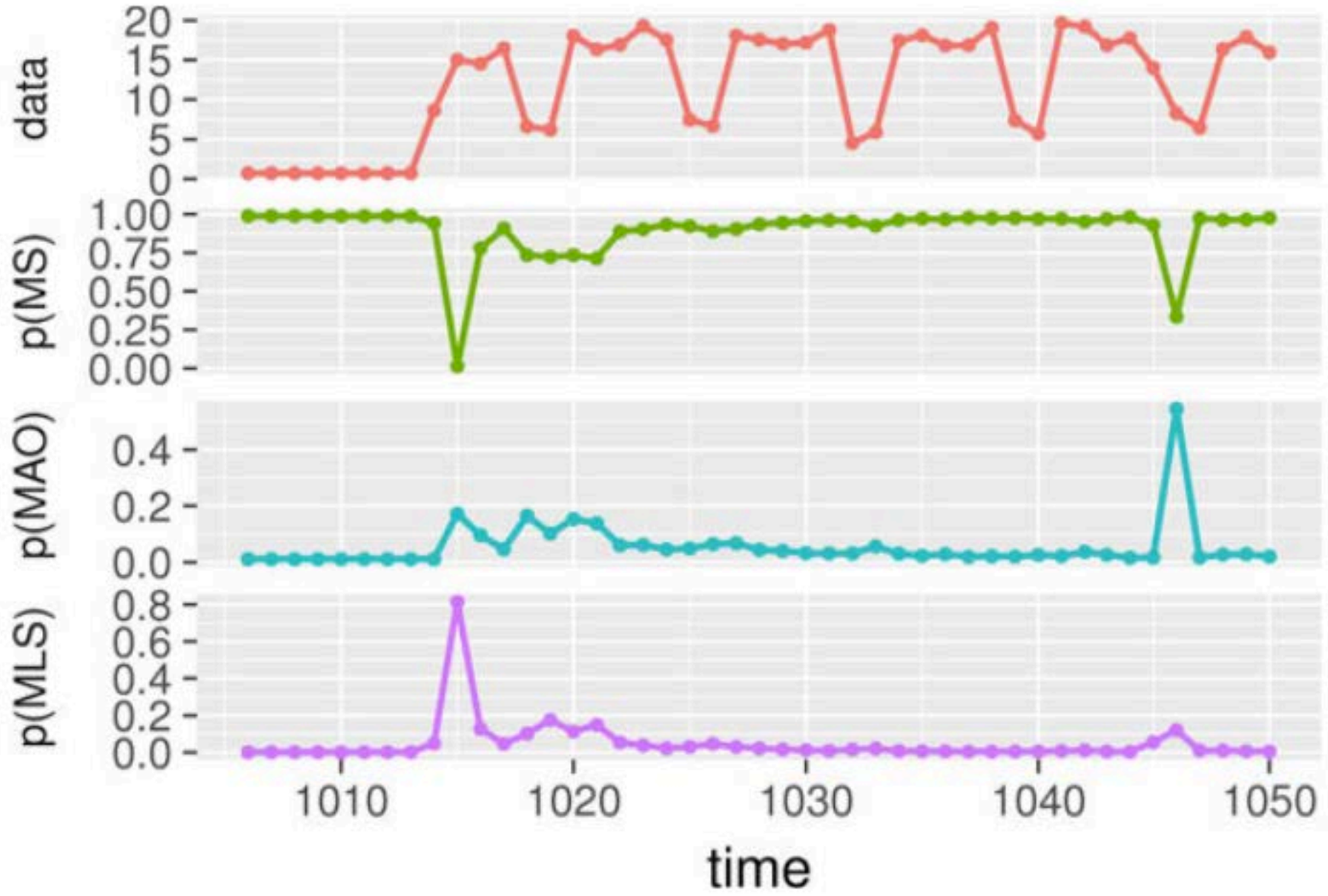


Fig. 1.

Applying the MPDLM method to a time series. The top graph shows the observations. The remaining graphs show the posterior probabilities of the three models (MS, MAO, MLS). There is a one-time-unit delay for the probability outputs.

Estimating DLM Parameters is Challenging

- No labeled data
- Use non-informative priors for different behaviors (even though MS is probably most common)
- Hypothesize hyper-parameters that estimate *V* and *W* for the different models
- Evaluated on both real data and various simulations
  - Real: 14 rules with ≥ 1 change points (22 total)
- Delay vs. False Positive Rate; AMOC is area under that curve

Other Workflow Issues

- Alerting
  - Escalation of alerts on non-response
  - BIDMC study of unread messages in Patient Portal (only ~3%)
- Importance of Communication
- Integration of all data sources
  - Failure of Google Health, Microsoft Health Vault, …

Lab Alerts

- Beth Israel experience, 1994
  - rising creatinine levels while taking nephrotoxic or renally excreted drugs
  - 21.6 hour reduction in reaction time
  - risk or renal impairment reduced to 0.45 of pre-intervention level
  - 44% of docs found them helpful, 28% found them annoying, 65% wanted them continued

Rind, D., Safran, C., Phillips, R. S., Wang, Q., Calkins, D. R., Delbanco, T. L., et al. (1994). Effect of computer-based alerts on the treatment and outcomes of hospitalized patients. Archives of Internal Medicine, 154(13), 1511–1517.

The communication space

- is the largest part of the health system's information space
- contains a substantial proportion of the health system information 'pathology'
- is largely ignored in our informatics thinking
- is where most data is acquired and presented

How big is the communication space?

- Covell et al. (1985): 50% info requests are to colleagues, 26% personal notes
- Tang et al (1996): talk is 60% in clinic
- Coiera and Tombs (1996,1998): 100% of non-patient record information
- Safran et al. (1998): ~50% face to face, EMR ~10%, e/v-mail and paper remainder

What happens in the communication space?

- Wilson et al. (1995): communication errors commonest cause of in-hospital disability/death in 14,000 patient series
- Bhasale et al. (1998): contributes to ~50% adverse events in primary care
- Coiera and Tombs (1998): interrupt-driven workplace, poor systems and poor practice

No of call events (No of successful connections) categorised by subject and call type among 10 hospital staff

Subject and role	Page call		Telephone call		Length of observation (hours: minutes)	Total No of events
	Sent	Received	Made	Received		
7 (consultant)	0	0	0	0	2:55	0
2 (house officer)	0	0	0	0	2:59	0
1 (consultant)	0	0	1 (1)	0	3:15	1 (1)
6 (senior registrar)	0	0	2 (2)	0	2:05	2 (2)
9 (house officer)	3 (0)	3 (3)	6 (6)	0	2:41	12 (9)
8 (nurse)	4 (2)	0	4 (4)	5 (5)	2:09	13 (11)
10 (house officer)	0	2 (2)	11 (10)	0	2:55	13 (12)
5 (senior registrar)	0	4 (4)	10 (7)	0	3:39	14 (11)
3 (nurse)	1 (0)	2 (2)	13 (4)	1 (1)	3:23	17 (7)
4 (senior house officer)	1 (1)	10 (10)	9 (3)	4 (4)	3:39	24 (18)
Total	9 (3)	21 (21)	56 (37)	10 (10)	29:40	96 (71)

Coiera, E., & Tombs, V. (1998). Communication behaviours in a hospital setting: an observational study. BMJ (Clinical Research Ed), 316(7132), 673–676.

ER communication study

- Medical Subject #4
  - 3 hrs 15 min observation
  - 86% time in ‘talk’
  - 31% time taken up with 28 interruptions
  - 25% multi-tasking with 2 or more conversations
  - 87 % face to face, phone, pager
  - 13 % computer, forms, patient notes

Implications

- Clinicians already seem to receive too many messages resulting in:
  - interruption of tasks
  - fragmentation of time, potentially leading to inefficiency
  - potential for forgetting, resulting in errors

Communication options

- We can introduce new:
  - *Channels,* e.g., v-mail
  - *Types of message,* e.g., alert
  - *Communication policies,* e.g., prohibit sending an e-mail organisation-wide
  - *Communication services,* e.g., role-based call forwarding
  - *Agents* creating or receiving messages, e.g., web-bots for info retrieval
  - *Common ground* between agents, e.g., train team members
- Synchronous:
  - face to face, pager, phone
  - generate an interrupt to receiver
- Asynchronous:
  - post-it notes, e-mail, v-mail
  - receiver elects moment to read

Automated messages

- *Notification* - that an event has occurred:
  - *Alert* (push)- draws attention to an event determined to be important, e.g., abnormal test result, failure to act
  - *Retrieve* (pull) - return with requested data
  - *Acknowledgment* (push or pull) - that a request has been seen, read, or acted upon

Notification systems

- Channel:
  - typically asynchronous, e.g., e-mail, pager, fax
  - synchronous modes feasible
- Message:
  - existing messages, e.g., lab alerts
  - new messages, e.g., task acknowledgment

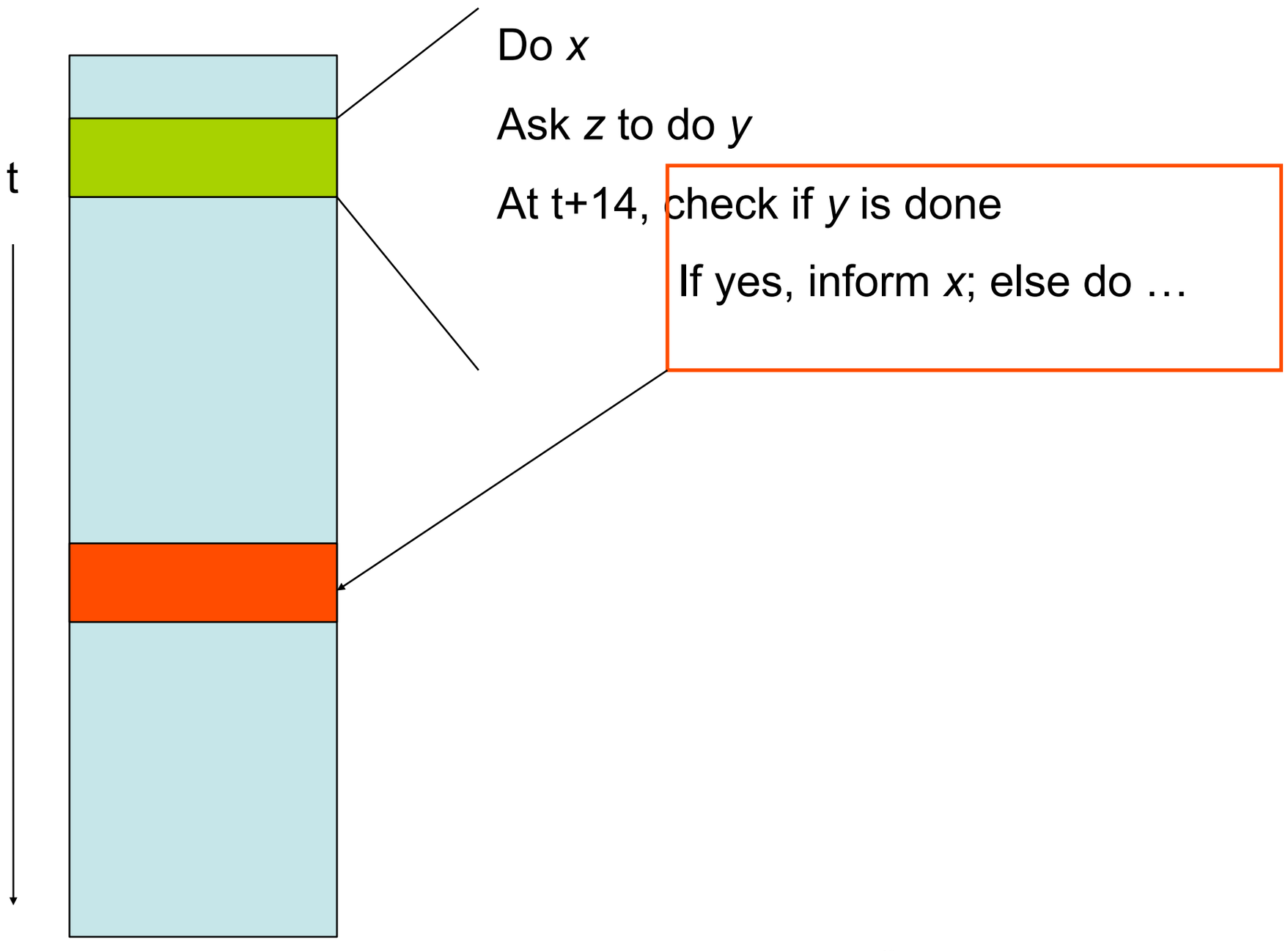
Effects of notification systems

- *Channel effect*: shift existing events from synchronous to asynchronous domain, reducing interruption
- *Message effect*: generate new types of events in the asynchronous domain, increasing message load, demanding time, and creating a filtering problem
- potential to either harm or help

How to keep from dropping the ball?

- Coordination
  - CSP, where some of the processes are people
  - Checking that others are “on track”
- Resource allocation
- Design of rational human-institution-technology systems

Workflow Engine
≈ discrete-event simulator



Google Health: A Personal Health Record

- In 2008, the service underwent a two-month pilot test with 1,600 patients of The Cleveland Clinic
- Starting on May 20, 2008, Google Health was released to the general public as a service in beta test stage
- 2011 Google announced it was retiring Google Health

- Partners: Allscripts, Anvita Health, The Beth Israel Deaconess Medical Center, Blue Cross Blue Shield of Massachusetts, The Cleveland Clinic, CVS Caremark, Drugs.com, Healthgrades, Longs Drugs, Medco Health Solutions, Quest Diagnostics, RxAmerica, and Walgreens
- Other than these partners, no facilities to enter data automatically
- No facilities at all to allow/encourage clinicians to look at these data
  - Missing integration with hospital/clinic EHRs

- Also see “Guardian Angel”, http://ga.org

6.S897 / HST.956 Machine Learning for Healthcare
Spring 2019