Predictive Modeling in Healthcare: Where We Are and What the Future Holds

Jonathan P. Weiner, DrPH

Professor of Health Policy & Management and of Health Informatics

The Johns Hopkins Bloomberg School of Public Health &The Johns Hopkins School of Medicine

jweiner@jhsph.edu, 410 955-5661

Keynote presented at the National Predictive Modeling Summit 9/22/08 – Washington DC

The topics I will cover:

- **Taking stock**
  - Current status of "predictive modeling" (PM) in health care in the US and globally
- **A shared understanding of the status quo**
  - Paradigms, frameworks and nomenclature
- **Selected R&D findings from Johns Hopkins**
- **Future directions**
  - Moving the state-of the-art forward in the e-health environment
  - Future frontiers and challenges

Some underlying reasons why the application of predictive modeling (PM) is increasing

- Health care needs are rising, resource availability is not. Tools like PM are one solution.

- Electronic health records and other Health IT (HIT) are expanding availability of inputs for models.

- As intensity of clinical and financial interventions increase, considering and adjusting for risk becomes critical.

- Care management / disease management (CM/DM) are facing many challenges. PM is viewed as essential to getting more value from these programs.

Some observations about the PM status quo

- Deriving “risk” information from diagnosis, pharmacy and prior-use data found in health insurance claims is now an accepted business practice in US health care.

- “Risk adjusted” rate setting in the commercial market and government-to-MCO risk-adjusted capitation rates are now the norm.

- Predictive Modeling is now standard for “risk identification” and “risk stratification within care management / disease management programs.

PM Status Quo - 2

- In most instances, there is limited transparency of the PM tools and their applications.
- Within the PM context, there has been limited interface between business, clinical, statistical and informatics disciplines.
- The top methodologies (when compared fairly) all have similar “predictive power.” Marketing hype aside, though some differences exist, there has been a high degree of commoditization.
- While there is growing consensus on methods, there is still lots of room for standardized applications and Impact evaluation.

Importance of risk factors in explaining health care use:

The underlying rationale for predictive modeling

Risk and Costs are Concentrated in a Small % of any Population

Distribution of Expenditures for US Medicare Enrollees (65+)			
	% of Medical Costs		
% of Enrollees	(FFS)	(MCO)	% of Rx Costs*
2%	24%	32%	11%
10%	60%	68%	36%
50%	96%	97%	91%

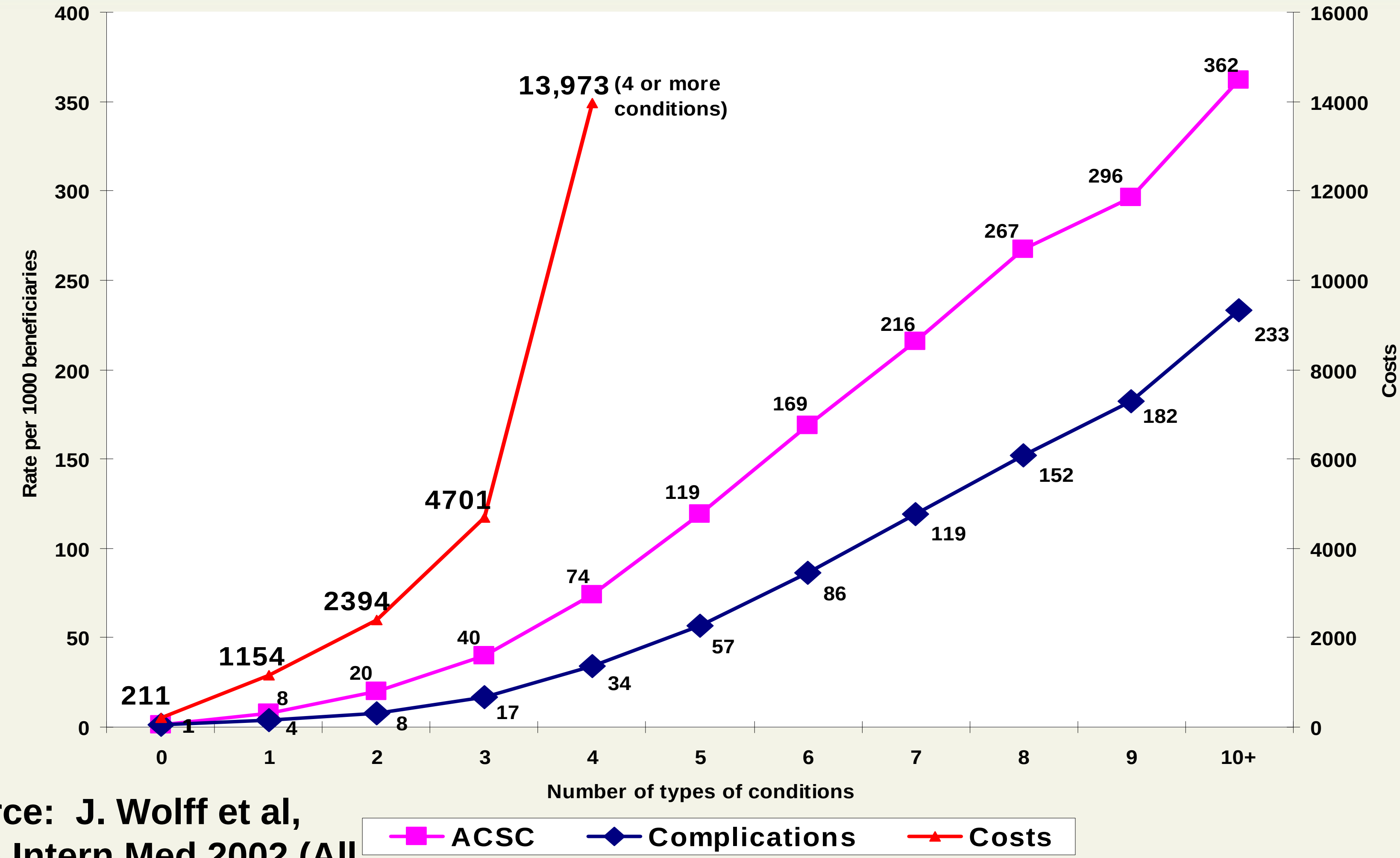
Sources of Data: FFS - '99 CMS 5% file. MCO- sample of 180,000 enrollees from several M+C plans in '00. * Based on prescription Rx claims from MCOs.

These patterns are linked to the prevalence of chronic co-morbidities (US Medicare 65+)

# Chronic Co-morbidities	% Pop.	Relative Cost (Per Pt.)	Est. % of Total Medicare Costs	Avg. # Unique MDs/Yr.	Avg. # Filled Rx / Yr.
5+	20%	3.2	66%	13.8	49
3-4	27%	.9	23%	7.3	26
0-2	53%	.1	11%	3.0	11

Data Source: G. Anderson et. al., Johns Hopkins Univ. 2003. (Derived from Medicare claims and beneficiary survey.)

The relationship between co-morbidity, hospitalization, avoidable events, and costs* (Americans 65+)



Source: J. Wolff et al, Arch Intern Med 2002 (All results are concurrent. Uses JHU ADGs)

Co-morbidity, rather than type of illness appears to be predictor of resource use

Expected Resource Use (Relative to Adult Population Average) by Level of Co-Morbidity* and Condition Type, British Columbia,

Type of Condition During Period	None	Low	Medium	High	Very High
Acute conditions only	0.1	0.4	1.2	3.3	9.5
1+ Chronic condition	0.2	0.5	1.3	3.5	9.8
1+ High impact chronic condition	0.2	0.5	1.3	3.6	9.9

***Co-Morbidity Levels based on Johns Hopkins ACG Morbidity Burden Bands. Source: Broemeling et al. Chronic Conditions and Co-morbidity among Residents of British Columbia. University of British Columbia, 2005.**

Paradigms, Definitions, Frameworks & Methods

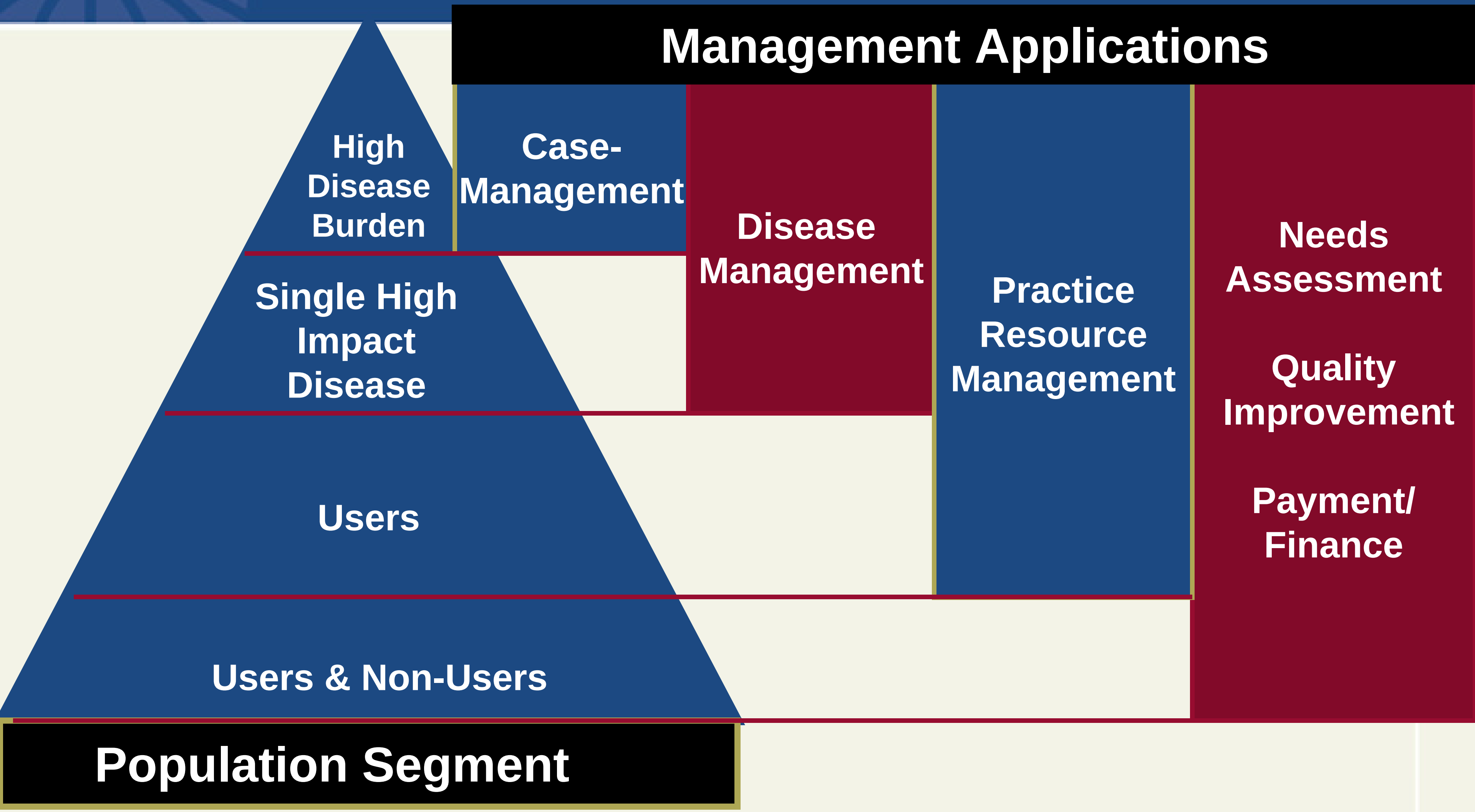
Working Definitions

- **Predictive modeling** is the prospective (or concurrent) application of person level risk measures and statistical analytic technique to identify individuals with high medical need who would likely benefit from care management interventions.

- **Risk adjustment** is the process by which the health status of a population is taken into consideration when:
  - setting budgets,capitation rates or premiums;
  - evaluating provider performance; or
  - assessing outcomes of care.

  (The terms predictive modeling & case-mix / risk adjustment are often used synonymously in these contexts).

The risk measurement pyramid



Predictive modeling / risk adjustment applications within health care

- **Financing, Payment, Planning**
  - Morbidity-adjusted capitation
  - Actuarial Rate / Premium setting
  - Allocation of budgets
  - Service targets
- **Provider Performance Assessment**
  - Profiling
  - Pay-for-Performance
- **Care Management**
  - Identification of high risk patients
    - Disease management
    - Case management
  - Population health monitoring
- **Quality**
  - Quality improvement
  - Quality monitoring
- **Research and Program Evaluation**

Key Components of a Typical PM Implementation for Care Management

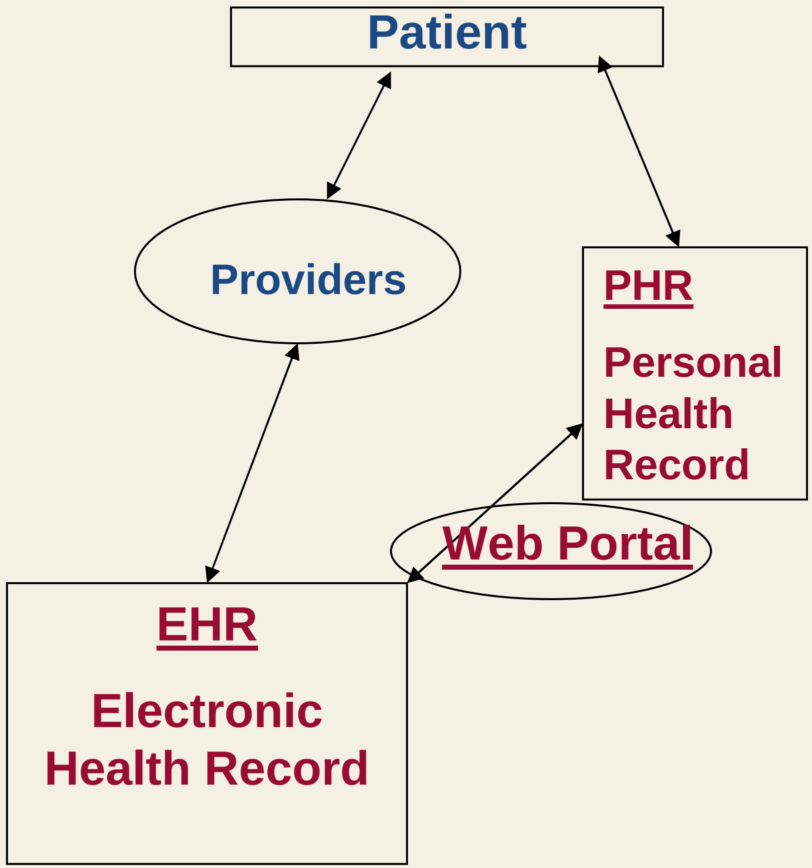
- **Collecting risk factor data**
  - **Administrative data sets, surveys, and “new” electronic sources**
- **Data warehouse/repository**
- **Analytic prediction model**
  - **Rules / Clinical Based (regression based, clinical trees)**
  - **AI / Data-mining (e.g. clustering, decision trees, neural nets)**
- **Reports / targeting information**
- **Care management interventions or other applications**

Data Sources for PM:
The Rapidly Changing Environment

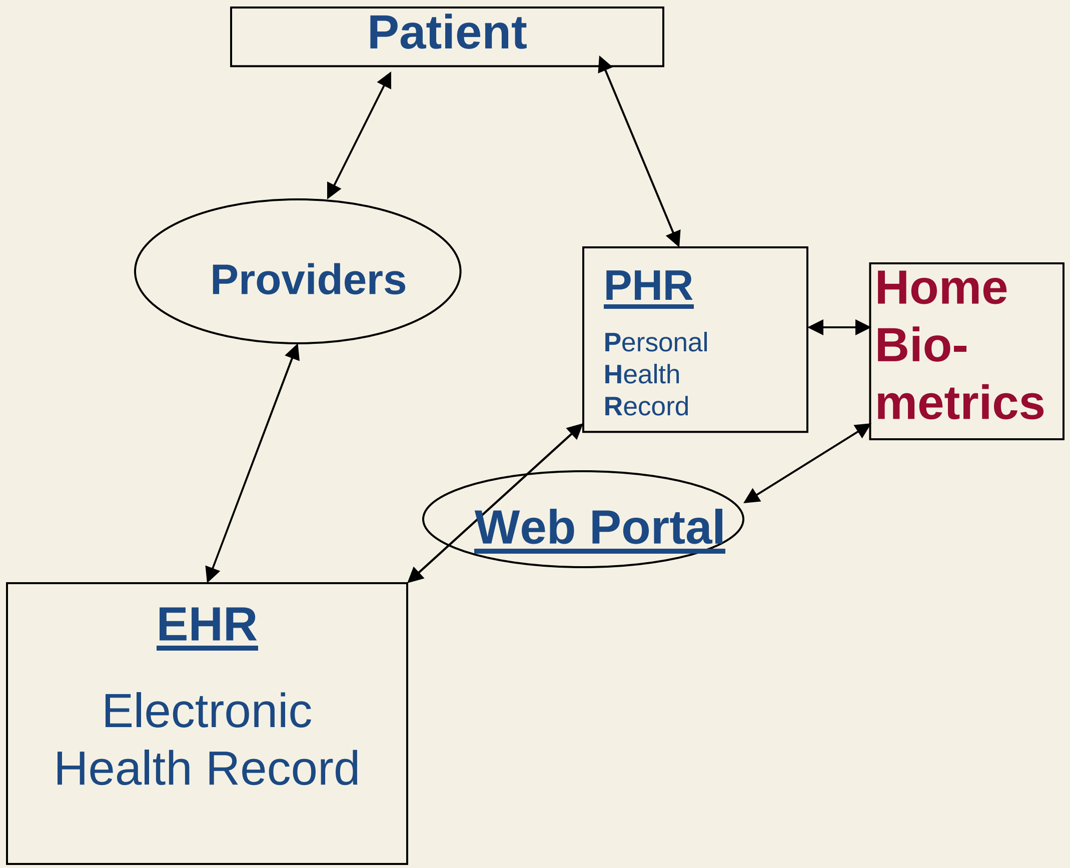
Type of measures for application to care management & assessment by data source

DATA SOURCE:	Denom.	Risk	Process	Outcome	Pt-Cent.	Cost
Electronic / Health IT						
Insurance files	X	X	X	X		X
EMR / EHR	X	X	X	X		X
PHR / web portal		X	X	X	X	X
Biometrics		X		X	X	
Non-electronic						
Paper medical record		X	X	X		
Surveys		X	X	x	X	

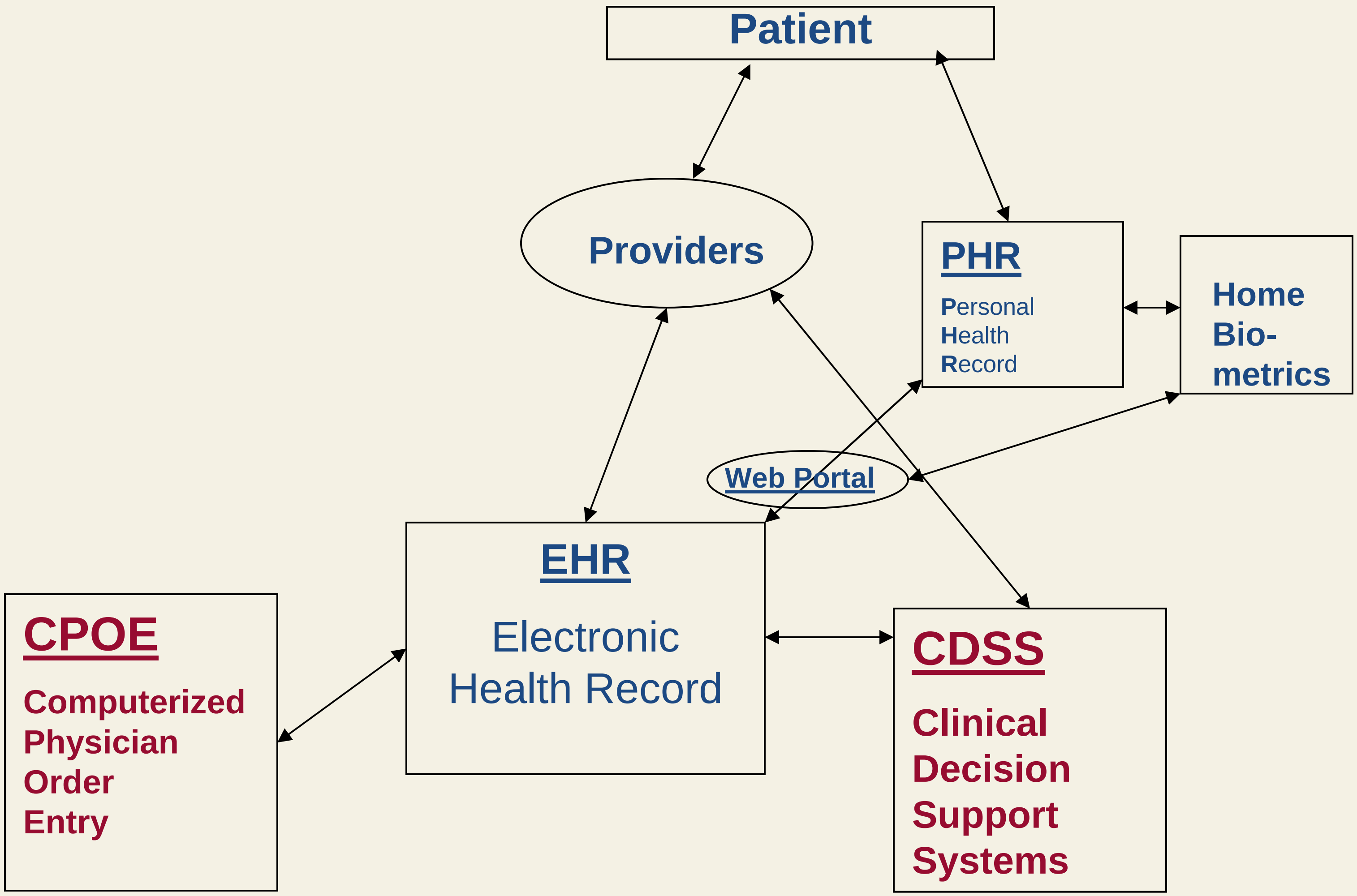
Electronic Health Records (EHRs) Health IT and the New “e-PM” context



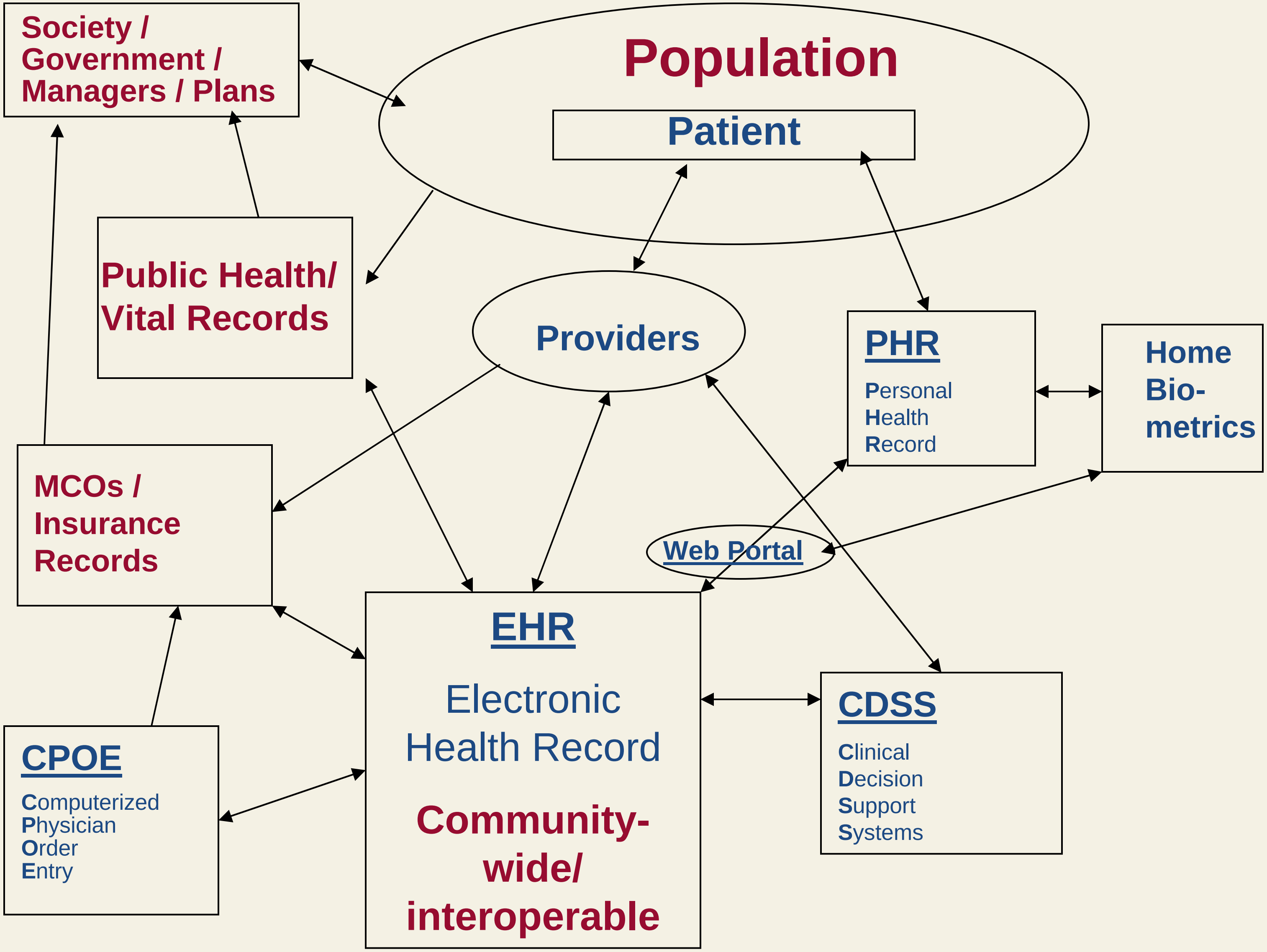
HIT Enabled Healthcare – The Next Frontier for PM - 1



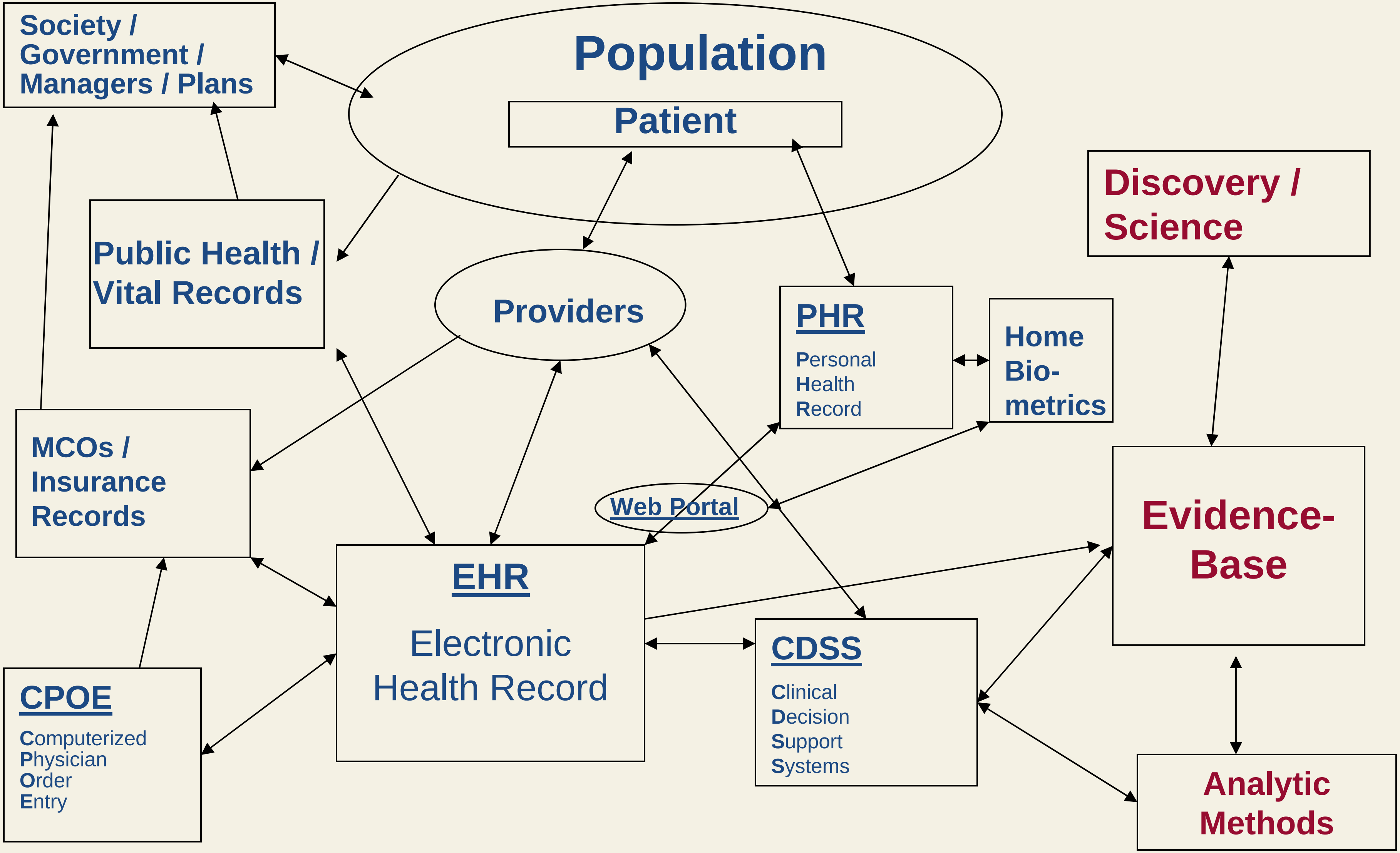
HIT Enabled Healthcare – The Next Frontier for PM - 2



HIT Enabled Healthcare – The Next Frontier for PM - 3



HIT Enabled Healthcare – The Next Frontier for PM - 4



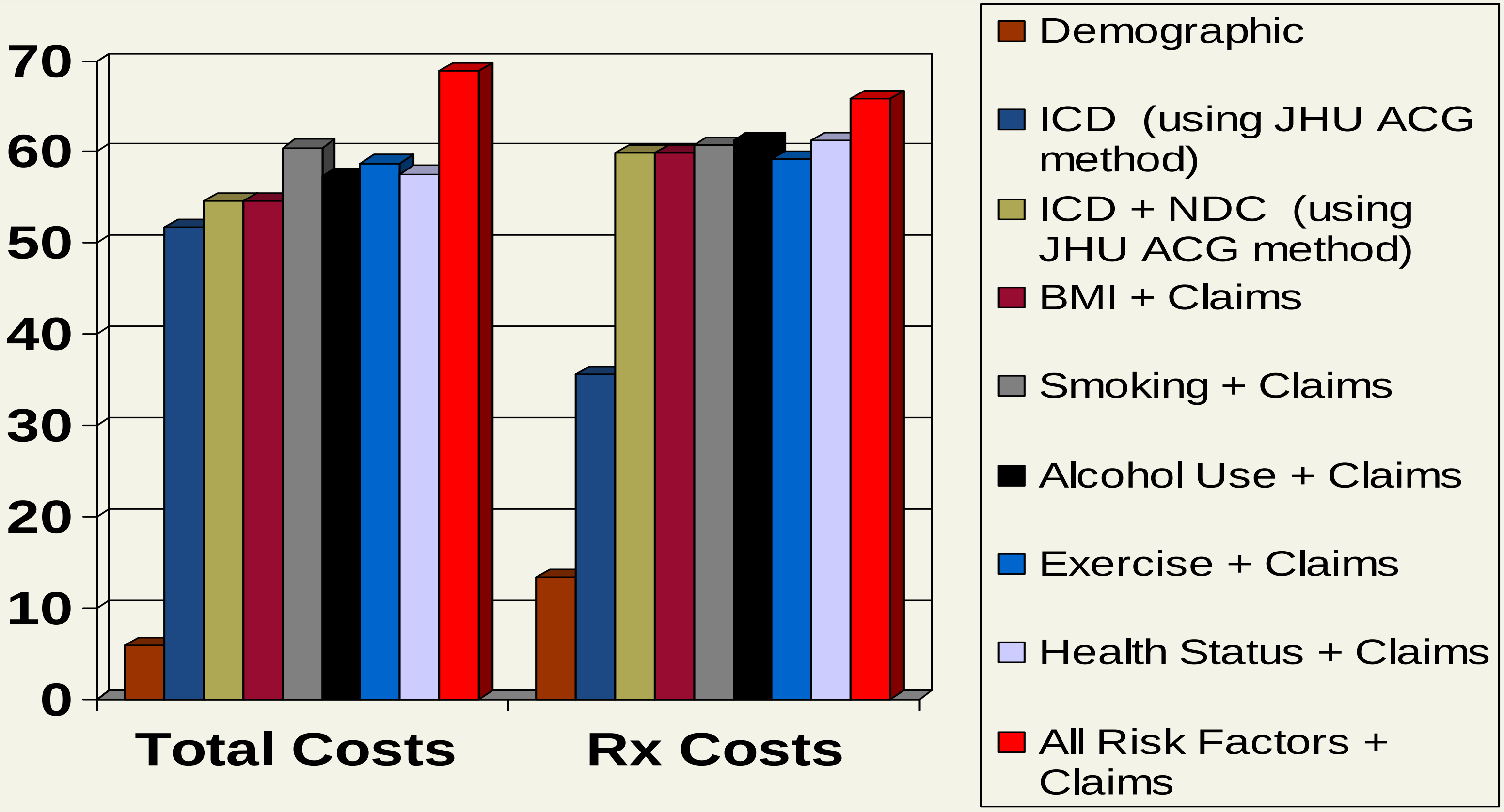
HIT Enabled Healthcare – The Next Frontier for PM- 5

In sum, there are many new potential electronic sources of risk factor input data for PM applications

- From EHR (e.g.,clinical findings, history)
- CPOE (e.g., e-prescribing and test ordering, lab & imaging results
- Home devices / biometrics
- PHRs / Pt. portal (e.g., preferences, function)
- Community surveillance / public health networks
- HIT usage/process (e.g., CDSS and workflow process)

Patient reports (via PHRs, web portals) potentially add important risk information for PM

Patient reported behavioral / risk factors potentially adds information (Figures reflect concurrent R-squared for cost)



Note. : Analysis by JHU using commercial health plan HRAs linked to concurrent costs (no truncation) from paid claims. Max n approx 70K. N's vary for each column and some censoring is likely due to response bias.

A word about <u>genomics</u> and its implications for EHR based PM and clinical decision support systems (CDSS)

- **Highly automated. Some organizations like the Iceland Ministry of Health / deCODE and Children's Hospital of Philadelphia are adding genomic data from their entire population to their EHRs**

Genomics, the EHR, PM and CDSS.

- Characterizes the genetic component of an individual's actual or potential disease and defines genetic sub-types ("clusters")
- Define treatment pathways for each genetic cluster including specific medication order sets and clinical workflow guidelines
- Can predict predict future disease patterns, outcomes and health care need over life course. (Raises many ethical issues).
- Key to development of evidence base regarding best course of treatment and subsequent development of improved CDSS.

The “M” in PM: Modeling concepts, techniques and issues

A word about “rules sets” vs. “natural relationship” models

- PM applies statistical / forecasting technique to maximize future individual level explanatory power.
- “Rule set” based models develop, test and validate clinically cogent models on large populations and then customize them to specific sub-populations.
- “Data mining” techniques identify “natural relationships” via “unsupervised learning” for each population.
- To avoid groupings that make little sense or are over-fitted, often clinical logic is imposed during data mining process.
- Most applications “meet in the middle” of the rules / data mining continuum.

A word about assessing model "success"

- There are many factors that go into model assessment
  - Statistical properties
  - Cogency / Transparency
  - Usefulness / Feasibility
  - All of the above is context driven.

- R-squared performance "horse races" are misleading (or worse). The relationship between R-squared and accuracy of case finding or actuarial group rate setting is limited.

The outcome / target of the forecasting model

- **To date, most PM has focused on cost**:
  - Actual cost
  - Outlier categories (e.g., top 1%)
- **Has also targeted utilization events:**
  - Hospitalization
  - ER /ED use
  - Poly-pharmacy
- **Clinical events**
  - Death (see case study)
  - BMI (See case study)
  - Outcome of specific clinical intervention

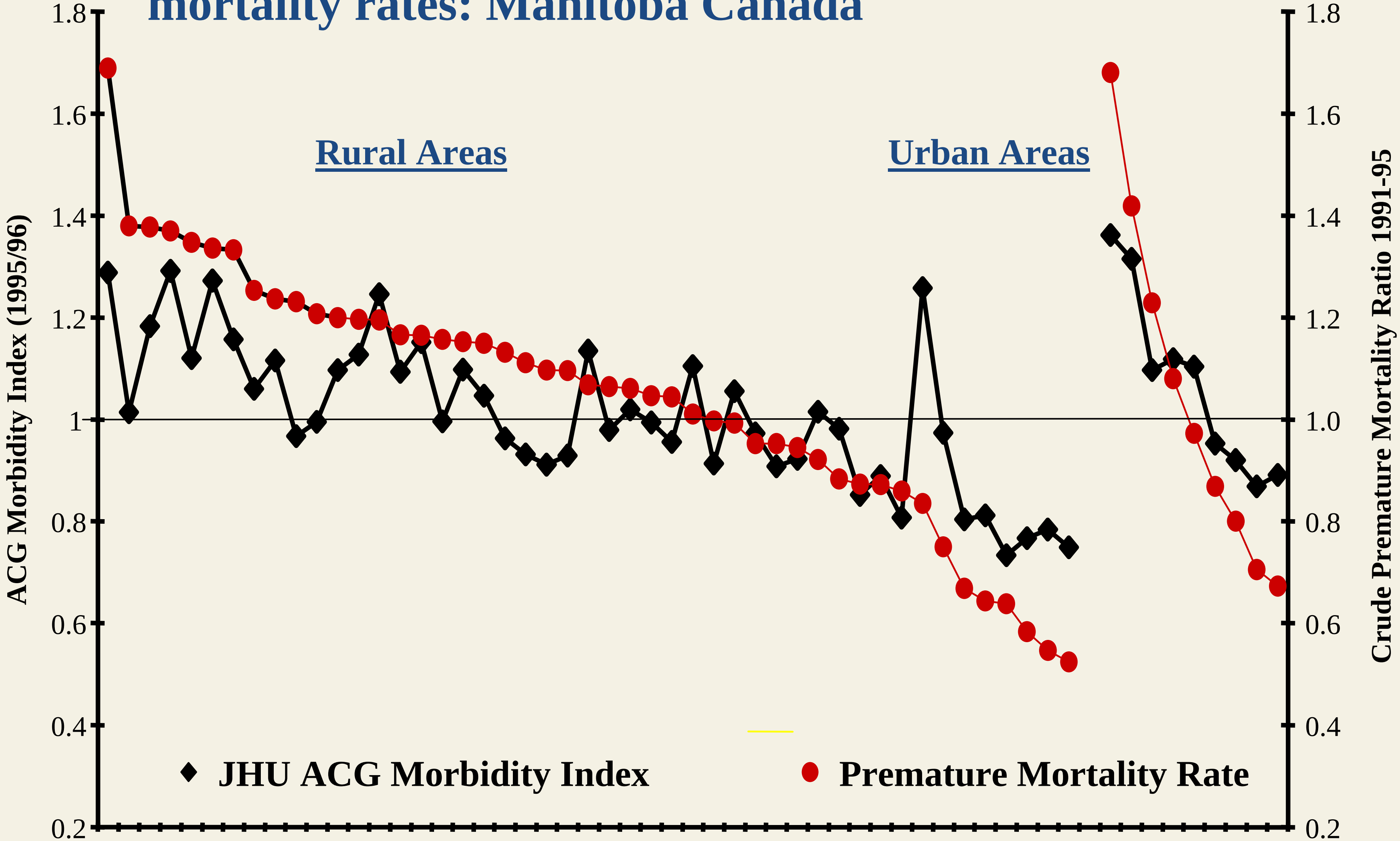
Using PM Risk Scores in Year-1 to Predict Clinical Outcomes in Year-Two for the entire Population of British Columbia Canada

Outcome	Highest Risk (Top 5% Risk Scores)	Lower Risk (Bottom 95% Risk Scores)	Risk Ratio (High / Low)
≥1 Hospitalization	27.0%	5.7%	5
≥1 ICU admit	1.9%	0.2%	8
≥1 CCU admit	2.0%	0.2%	10
Died	46 / 1,000 population	5 / 1,000 population	9

Overall Mortality Rate: 7 / 1,000 population

Source: British Columbia linked database (n=3,800,000). Using Johns Hopkins ACG predictive model.

Predictive modeling and population health status. ICD-based PM models in geographic areas parallels premature mortality rates: Manitoba Canada



(Source: Based on Manitoba Provincial Claims. See: R. Reid 2000)

Claims Based PM / Propensity Score to Identify Risk of Obesity

- Aim: To develop a predictive model / propensity score to identify obese health plan members using claims data
- Method -- Uses BMI from HRA survey and in-office exams to generate a predictive model or “propensity score” for BMI≥35 based on information in claims.
  - i.e., age, gender, presence of certain ICD, NDC codes
- Could be used by plans lacking BMI data for CM / DM programs or for population health needs assessment
- Data Source: 80,000 HRA respondents in 4 health plans, excluding bariatric surgery patients, and pregnant women

Obesity Predictive Model / Propensity Score: Results of Screening Application

Top x% of Sample	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
1	2.91	99.96	97.29	67.21
5	13.46	99.25	90.02	69.54
10	24.27	97.17	81.16	71.86
25	46.33	85.72	61.97	76.07
30	51.50	80.80	57.40	76.83
33	55.19	77.65	55.37	77.53
50	68.72	59.41	45.96	79.08

Based on Johns Hopkins ACG model using ICD and NDC risk information. RoC / AuC of Model is .72 .

(Study by J. Clark et al, JHU)

Wish list for “targets” of future PM applications

Survival
- Longevity / Death

Anatomic-Physiologic
- Injury
- Disease development
- Disease complications
- Disease severity/stability
- Physiological stability
- Biological adaptability
- Growth
- Biological risk

Behavior
- Engagement
- Coping
- Behavioral risk

Functioning
- Mobility
- Self management
- Communication
- Interpersonal relationships
- Intellectual functioning

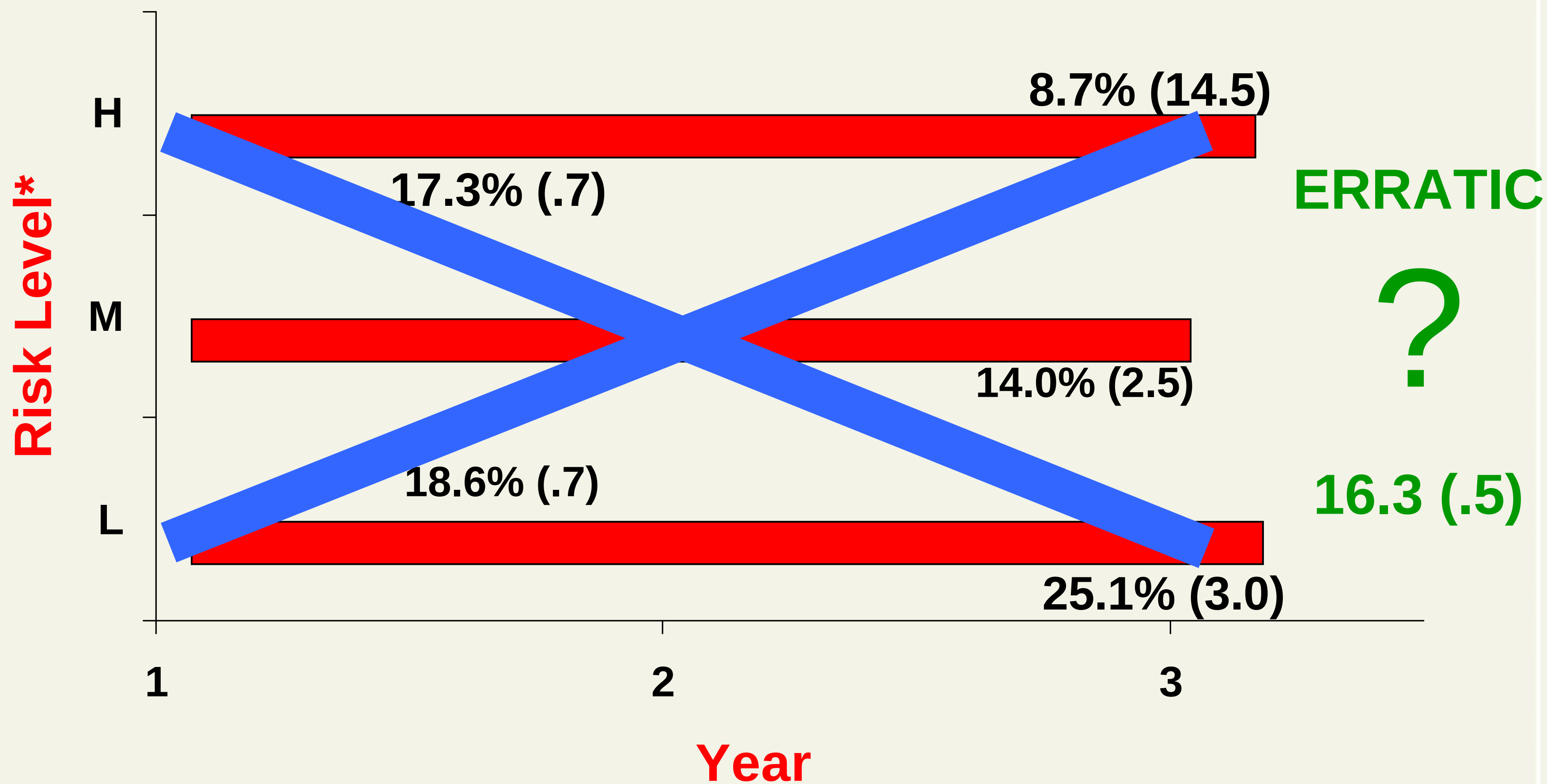
Perceptions
- Interpersonal connectedness
- Symptoms
- Comfort
- Well-being

Some PM Development Priorities

- Embracing new digital data sources or **"e-PM".**
- Better targeting of patients where future high-risk events are "avoidable" or "amenable" to change.
- Integrating risk measurement / PM and quality improvement; particularly related to care coordination across providers and settings.
- Developing alternative approaches for episode assessment that capture multi-morbidity, complex nature of care patterns while retaining holistic perspective.

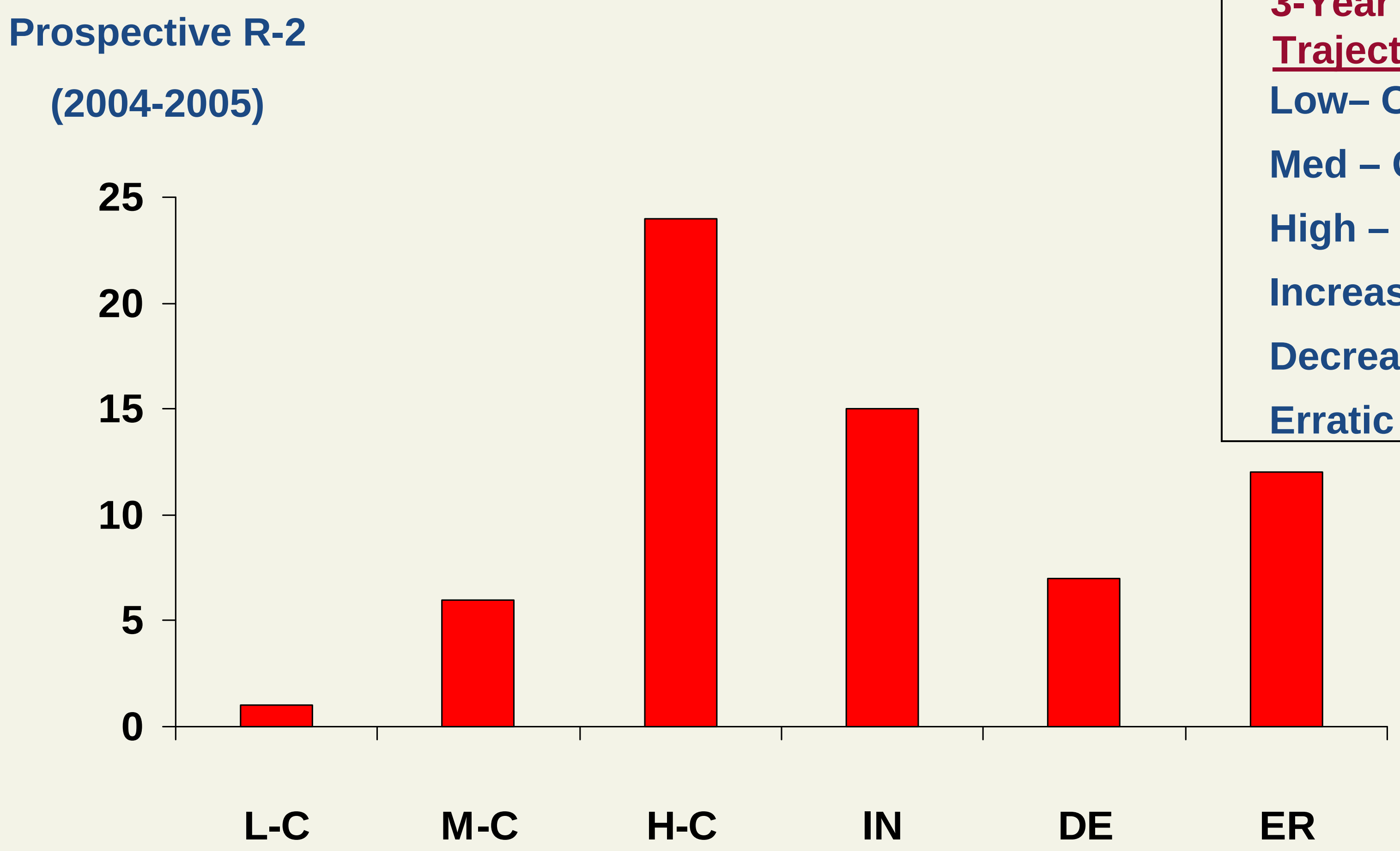
One important R&D area is understanding the time-trajectory of individuals and populations over time

Multi -Year “Trajectory” of Care May Offer New Information on Risk **(Figures reflects % of population in category & odds ratio of actual to expected incidence and national sample from Taiwan)**



*** Based on Johns Hopkins ACG Morbidity Burden Bands. Source: Taiwan’s National Health Plan 2002-2004. Hsien-Yen Chang, JHU Doctoral Thesis**

Prospective Explanatory Power Varies Considerably for Different “Trajectory Groups”



3-Year (2002-2004) Trajectory Groups

Low– Constant

Med – Constant

High – Constant

Increasing

Decreasing

Erratic

Model based on 2004 ICD based ACG risk markers to predict 2005 untruncated total cost. Source: Taiwan’s National Health Plan. Hsien-Yen Chang JHU Thesis

Some Longer-Term PM “Frontiers”

- Integration of PM with real-time clinical decision support systems and electronic health records.
- Moving from “black box” empirical models towards those based on biologic, humanistic and system “mechanistic” models of cause and effect.
- Integrating “PM” into population level decision support.
- Finding ways to support the learning health care organizations and development of global evidence-base.

Predictive Modeling in 5+ years: e-PM = CDSS?

- Will the predictive modeling and clinical decision support (CDSS) fields overlap -- or merge -- in the future?

- Today PM tends to be administrative and periodic and CDSS tends to be ongoing and real-time. PM is likely to become more like CDSS.

Clinical Decision Support and PM moving forward

- A current working definition of Clinical Decision Support Systems (CDSS) is as follows:

  CDSS is an e-health tool that applies:
  - an analytic model;
  - available patient risk factor data; and
  - the current knowledge-base

  to influence the choices of clinicians, consumers and other decision makers in order to improve health care value.

Will this be the definition of PM in the future?

Some of the future R&D challenges as CDSS and PM application become more integrated

- Knowledge / evidence acquisition
- Model-based reasoning
- E-health system integration with the clinical environment
- CDSS systems in support of more complex decisions
- Population based applications
- Integration with value / cost / coverage decisions
- Personal preferences
- Ethical issues

In Sum, Future Predictive Models Will (or Should) …

- Become part of the electronic health care workflow
- Be more finely tuned to specific individuals and populations
- Predict health outcomes beyond cost
- Target broader timeframes
- Be more accurate
- Involve more complex modeling
- Become more transparent and less hyped
- Be applied by a wider array of end-users
- Keep all of busy for the rest of our careers!!

In closing

“ Predictions are hard, especially about the future.”

Niels Bohr
Nobel Laureate in Physics

Contacts / Acknowledgements

ACG Web Site:

- www.acg.jhsph.edu

Contacts:

- **Amy Salls – DST Health Solutions Inc. (Distributor of ACGs in US & Canada)**
  (508) 405-0297 **asalls@dsthealthsolutions.com**
- **Dr. Karen Kinder-Siemens, Director International ACG Operations (Germany) (kkinder@jhsph.edu)**
- **Professor Jonathan Weiner**
  (jweiner@jhsph.edu)

Acknowledgements:

- **I would like to thank my colleagues at Johns Hopkins and at other institutions for allowing me to share their research findings and other material that I have adapted for this talk.**