Natural Language Processing:
Applications for Enhancing Clinical Decision-Making

April 24, 2012

Frank A Sonnenberg, MD, FACP, FACMI
Professor of Medicine
Medical Director of Clinical Information Systems
UMDNJ Robert Wood Johnson Medical School

UMDNJ
ROBERT WOOD JOHNSON MEDICAL SCHOOL
University of Medicine & Dentistry of New Jersey

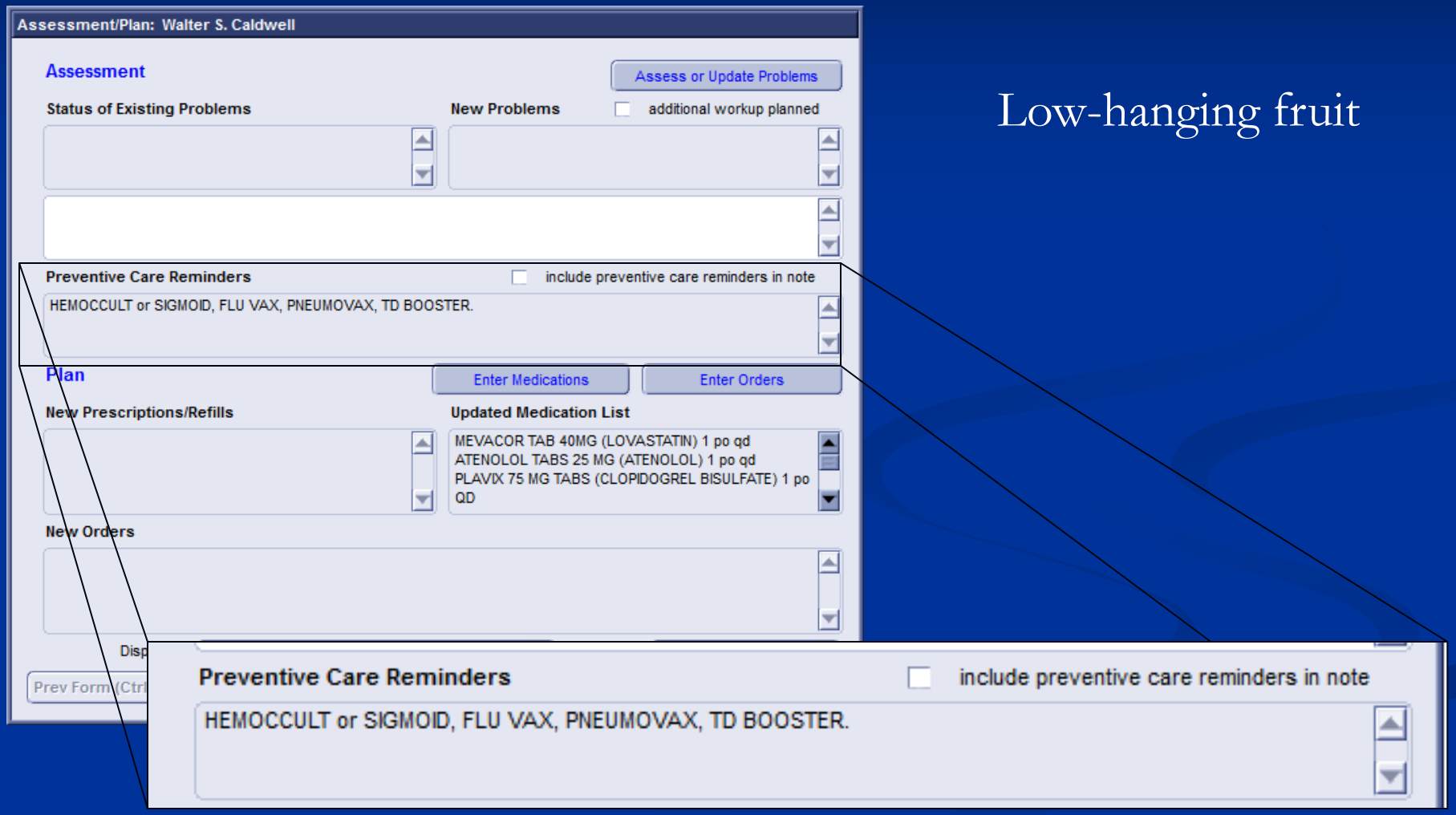
CDS consortium

My Perspective

- General internist for 30 years
- Medical Director of Clinical Information Systems
- Co-investigator, Clinical Decision Support Consortium (CDSC)

Clinical decision support is not readily available and not easy to implement with currently available tools.

Typical Decision Support



Low-hanging fruit

Current Challenges

- Current generation EHR's are not designed to optimally organize patient information – still *heavily document and episode-based.*
- Historical facts not readily accessible
- Certain kinds of data are not captured well as discrete data.
  - Symptoms
  - Physical findings
  - Previous procedures
  - Measurements supporting diagnoses

Instantiating variables to support decision rules

- Supplying variables automatically from the EMR requires capturing them during routine clinical care.
- Pertinent variables must be identified.
- EMRs, in general do not currently employ standardized terminology.
- Much critical information is not captured in EMR variables.

Two example guidelines

- The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7)

  JAMA. 2003;289:2560-2571

- Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)

  JAMA. 2001;285:2486-2497.

Variable types

- Simple – single observations
  - e.g. "systolic blood pressure"
- Calculated –not observed directly, but calculated from other observations
  - e.g. "Age" calculated from birth date
  - e.g. "Body mass index" calculated from height and weight.
- Complex – variables defined in terms of other variables or observations
  - e.g. "elevated coronary risk"
  - e.g. "metabolic syndrome"

- Example: "metabolic syndrome" – defined as any 3 of
  - Abdominal obesity – in turn defined in terms of abdominal circumference
  - Elevated Serum triglycerides
  - Low HDL cholesterol
  - Elevated blood pressure
  - Elevated fasting blood glucose

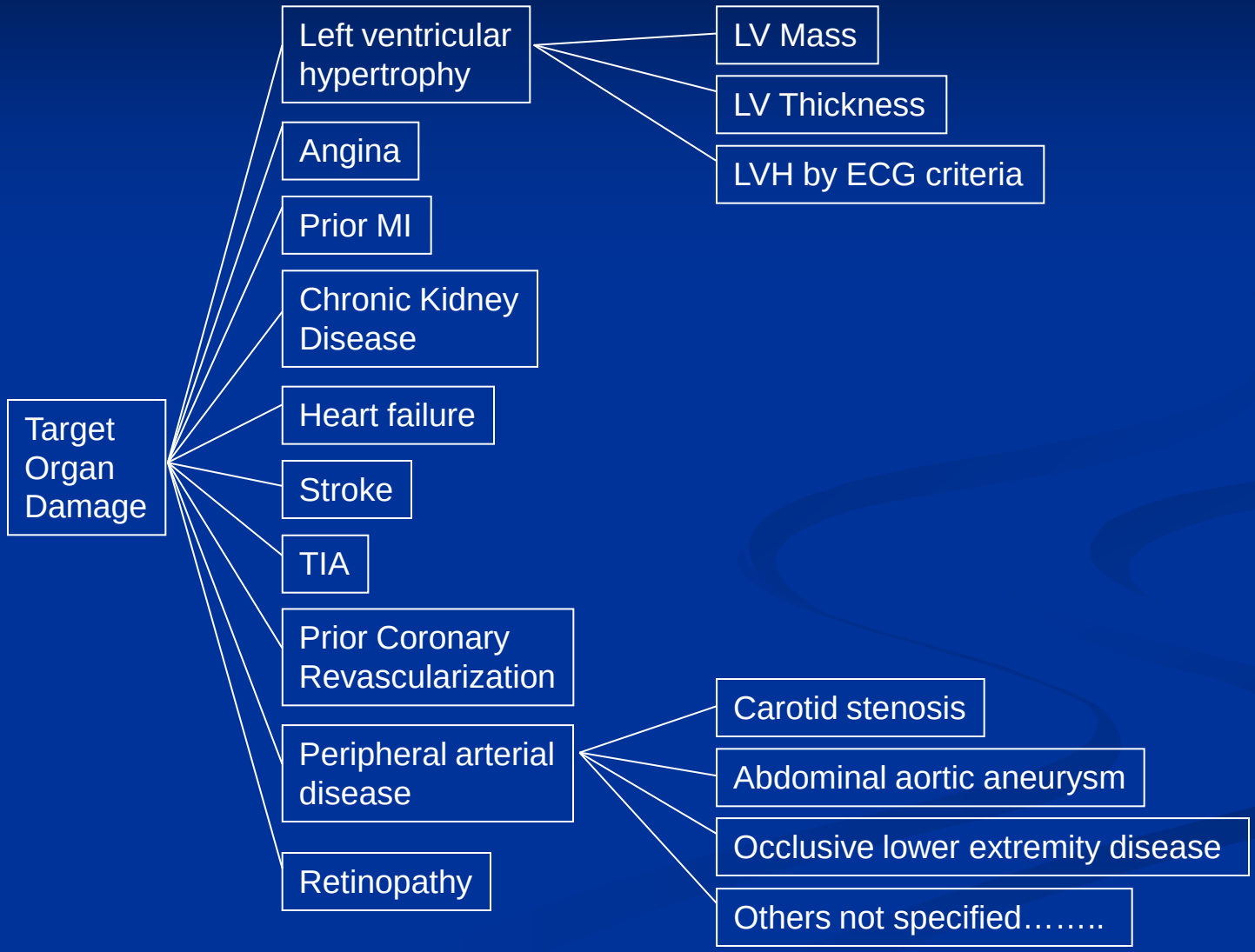
Variable Types Observed

Category	Subcategory	Number	Percent
Simple variable	Direct observation	42	25%
	Health issue	40	24%
	Medication	4	2%
	Need to ask clinician	15	9%
Calculated		32	19%
Complex term		36	21%
Total distinct terms		169	100%

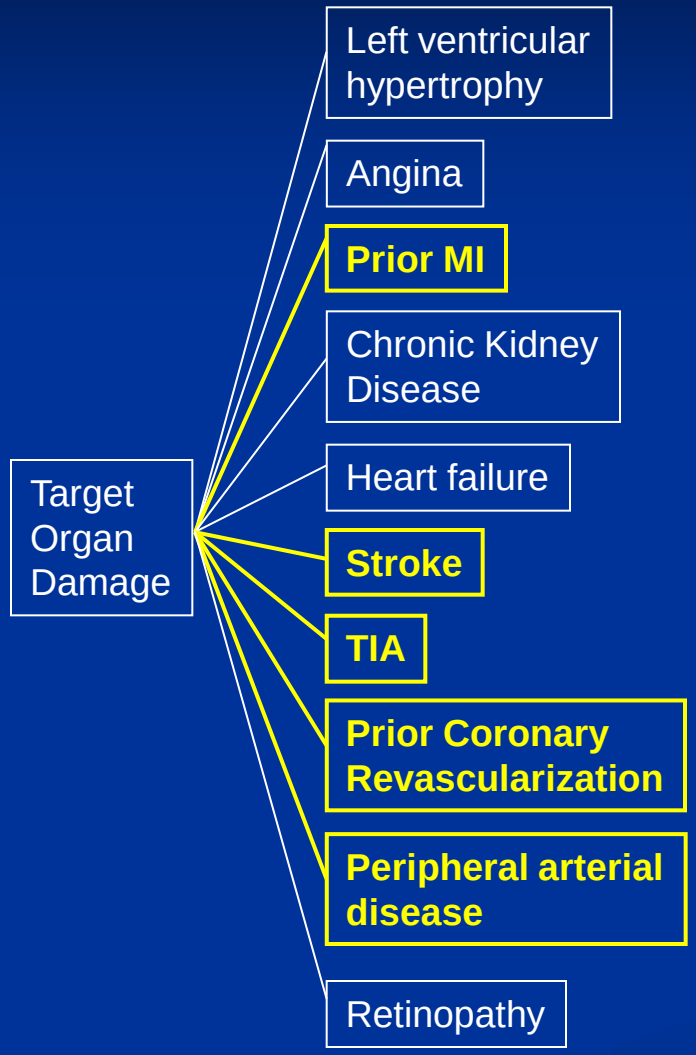
Findings

- Only 51% of variables were simple terms collected in the form required for application of the guideline.
- 40% of variables depend on other simple terms or required calculations
- There were many undefined terms: e.g. “children”, “adolescents”, “elderly”, “end-stage heart disease”
- Some terms were defined only outside the guideline e.g. “elevated coronary risk”.

Variable Dependencies

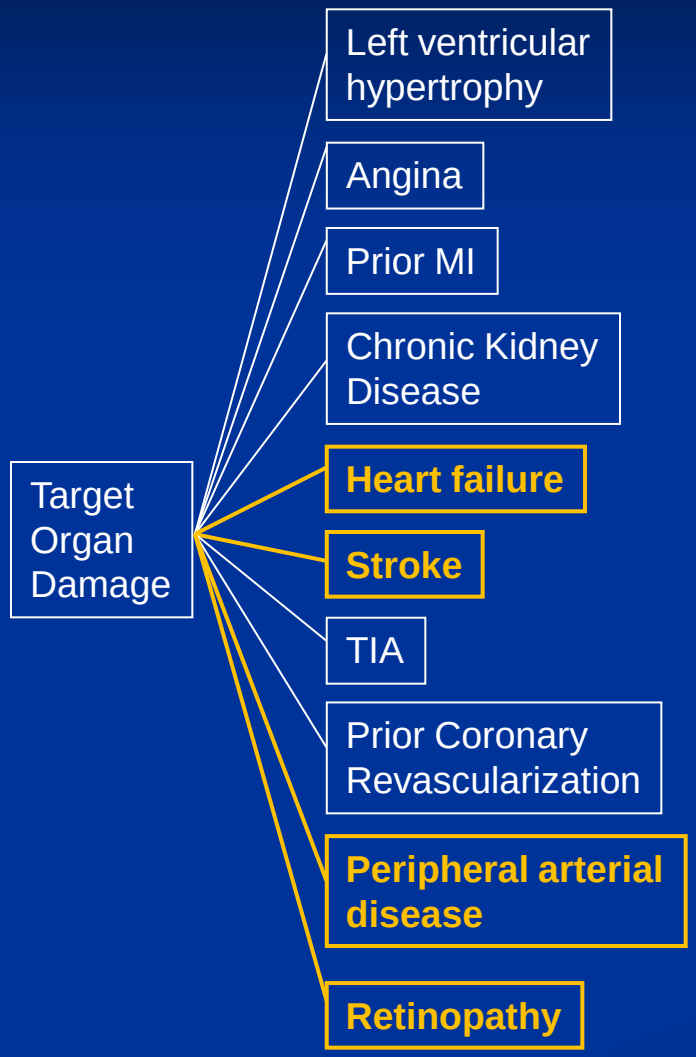


Variable Dependencies



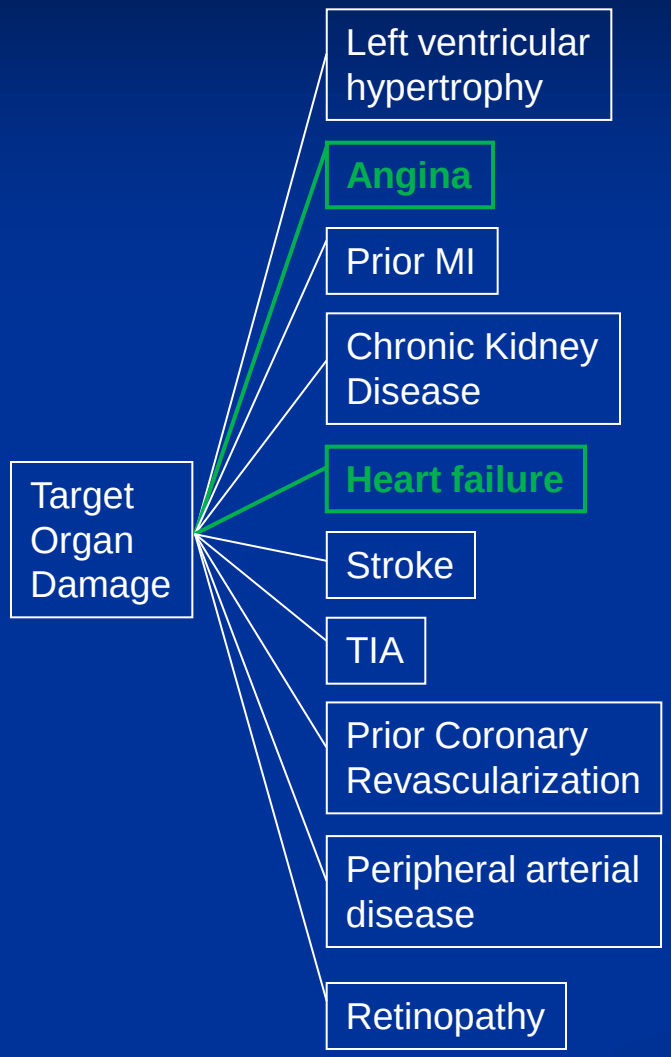
Dependent on comprehensive history

Variable Dependencies



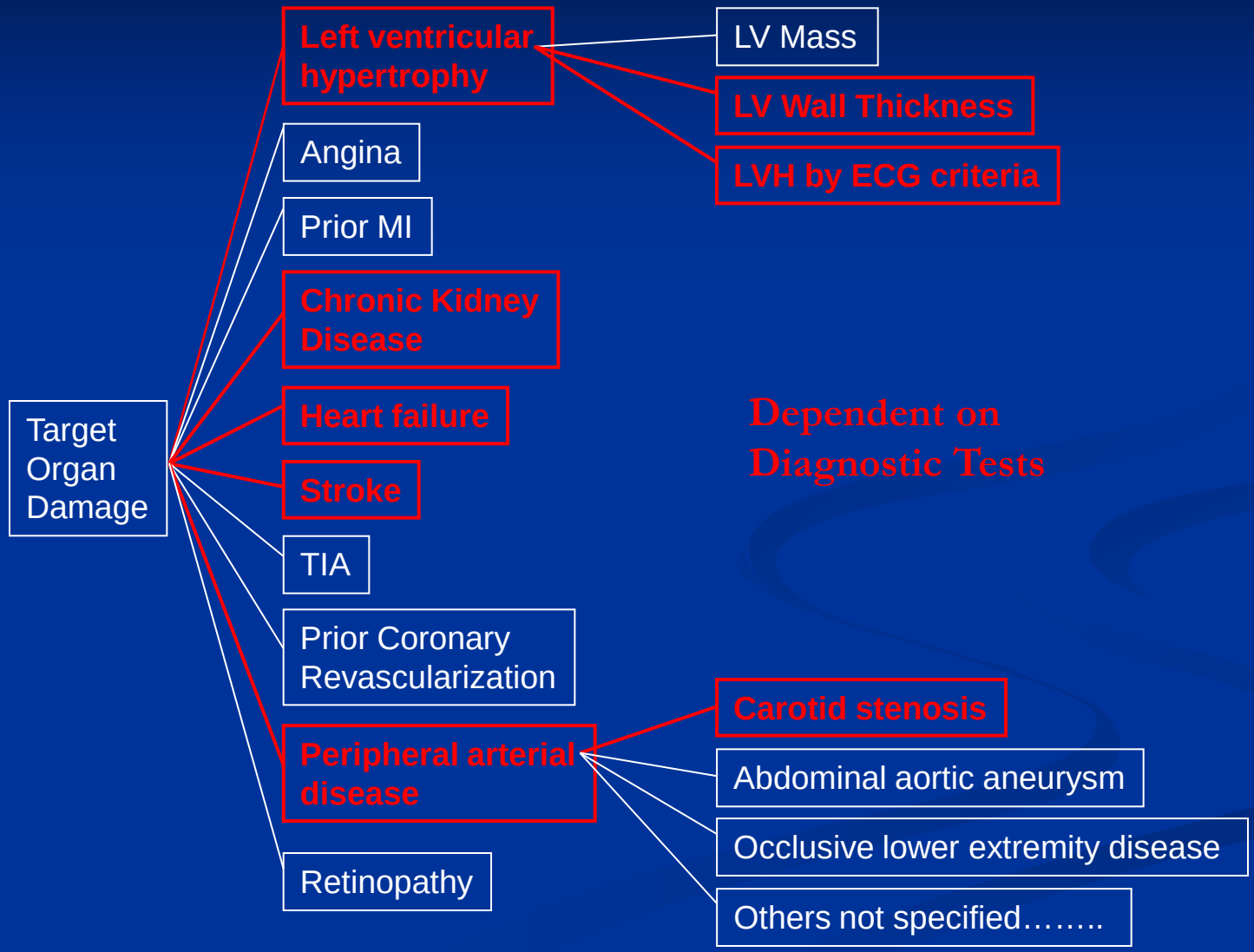
Dependent on Physical Exam

Variable Dependencies



Dependent on Symptoms

Variable Dependencies



LVH and CHF

From text-based Echocardiogram report

Tech:
Test Indications: CHF
Echo/Color Flow: X Doppler: X
M-Mode Measurements Technically Adequate: X Yes

		(Normal)			(Normal)
LVEDD	5.2	(3.5-5.7 cm)	LVESD	4.1	(2.3-3.9 cm)
LVPW Thick	1.0	(0.6-1.1 cm)	IVS Thick	1.0	(0.6-1.1 cm)
RVEDD	1.2	(0.7-2.6 cm)	LA Size	3.0	(1.0-4.0 cm)
AO Root	2.8	(2.0-3.7 cm)	ACS	1.8	(1.5-2.6 cm)

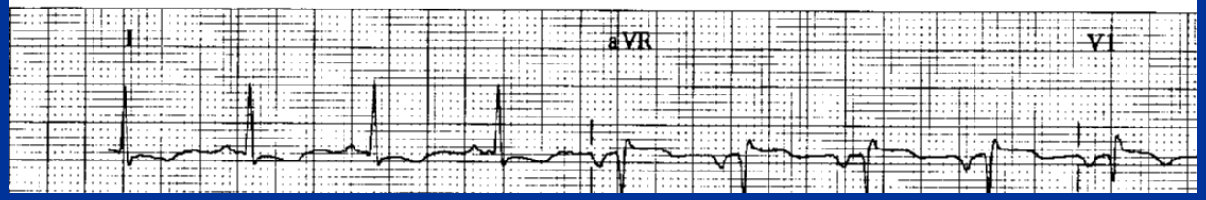
Two Dimensional Imaging:
2D echocardiographic views of the left ventricle show a normal size left ventricular cavity with segmental wall motion abnormalities. The basal septum and anterior wall are akinetic. The apex and inferior wall are hypokinetic. Overall ejection fraction is between 30-35%. There is normal biatrial size. The aortic root is normal.

From EKG report

Rate	94
PR	145
QRSD	82
QT	359
QTc	449
--AXIS--	
P	74
QRS	-17
T	124

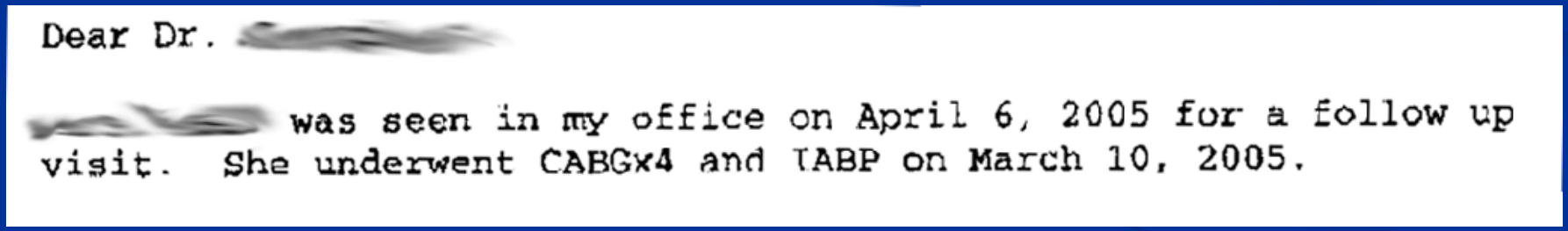
. Normal sinus rhythm, rate 94
. Leftward axis, probably normal for age
. Right atrial enlargement
. Early transition
. Consider left atrial enlargement
. Anterolateral ST-T abnormalities
. . Consistent with ischemia

- ABNORMAL ECG -



Evidence of coronary revascularization

In a letter from a consultant:



Dear Dr. [redacted]

[redacted] was seen in my office on April 6, 2005 for a follow up visit. She underwent CABGx4 and IABP on March 10, 2005.

Evidence of peripheral vascular disease

From carotid duplex report

Impression:

80-95% stenosis of the left internal carotid artery.
60-79% stenosis of the right internal carotid artery.
Mild plaquing of the common carotid arteries bilaterally.
Severe plaquing of the left carotid bulb and internal carotid artery.
Moderate plaquing of the right carotid bulb and internal carotid artery.
Mild plaquing of the external carotid arteries bilaterally.
Normal bilateral vertebral flow.

From JNC-VII Hypertension guideline

for HF.[40] In asymptomatic individuals with demonstrable ventricular dysfunction, ACE inhibitors and β-blockers are recommended.[52,62] For those with symptomatic ventricular dysfunction or end-stage heart disease, ACE inhibitors, β-blockers, ARBs, and aldosterone blockers are recommended along with loop diuretics.[40,41-48]

Documentation of symptoms

In a letter from a consultant:

Dear Dr. _____

I had the pleasure of seeing Ms. _____ in the office today for continued cardiovascular care after her March 2005 four-vessel coronary artery bypass surgery.

Overall, she is doing well without any shortness of breath or edema. Her physical activity now includes ~~normal activities of~~ daily living as well as walking around department stores. She is tolerating her medications and remains nonischemic.

Conclusions

- Application of decision rules requires instantiation of a large number of clinical variables.
- Many of these variables are not captured as discrete data, even in EHR's.
- Many important data items are available only as textual entries in narrative reports.
- Current decision support systems cannot make use of all the information that is present in EHR's.
- For the foreseeable future, natural language processing will be the only way of capturing these data from the EHR.