Data Quality
SHARPn

Aug. 2013

Recent summary of goals

- **Objectives**
- 1. Enumeration of data sources for each of 4 types of data:
  - a) Diagnoses
  - b) Laboratory values
  - c) Vital signs (Ht, Wt, BMI, SBP, DBP, HR)
  - d) Medications
- 2. Characterize sources, availability, quality, other characteristics
- 3. Draw samples from each source
- 4. Compare sources and data
  - a) Within institution
  - b) Across institution

Four to five manuscripts

- “John Henry” study. Compare machine and person in screening patients
- Intra-institution heterogeneity
  - Parallel analyses of data variation within site
- Inter-institution heterogeneity
- Data mining to enhance space of relevant terms
- BMI in the EHR at 2 sites

1) John Henry Project

Research and applications

Comparison of Approaches: Manual Chart Reviews vs. Use of Electronic Algorithms as a Supplement to Manual Chart Review in Screening for Eligibility for a Clinical Intervention

Gary M Lubin MBA, CPA, CGMA, CITP[1], Jeff Tarlowe MBA[1], Kent R Bailey PhD[2], Christina M Wood-Wentz MS[2], Wendy Gay RN, BAN[2], Deborah Olson RN, MSN[2]

Author Affiliations
[1]Centerphase Solutions, Inc., NJ [2]Mayo Clinic, MN

Charts requiring manual review to find 15 patients

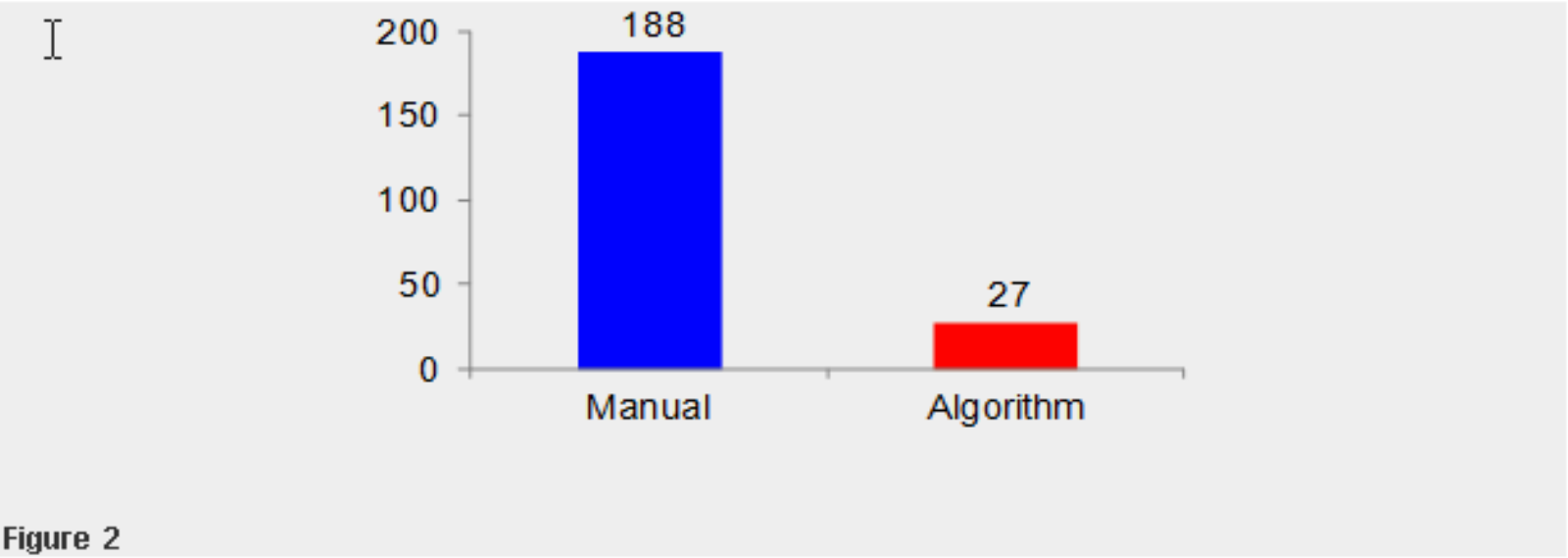


Figure 2

Cost comparison

Prescreening 15 Patients For A Clinical Study

Patients (charts	*Manual*	*Algorithm*
Total "At Risk" Patients	*15*	*15*
Charts Reviewed	*188*	*27*

	First Study		*Each Additional Study*	
	Manual	*Algorithm*	*Manual*	*Algorithm*
Time (minutes)				
IT Analyst time to implement algorithm	-	*492*	-	-
Clinical Nurse Specialist chart review time	*671*	*121*	*671*	*121*
Incremental time due to false positive	*30*	-	*30*	-
Clinical Nurse Specialist preparation time	*90*	*90*	-	-
Total Time:	*791*	*703*	*701*	*121*
Cost (dollars)				
IT Analyst cost for algorithm	-	*$473.36*	-	
Clinical Nurse Specialist cost for chart reviews	*$755.90*	*$209.59*	*$666.51*	*$120.19*
Cost of incremental time due to false positive	*$29.80*	-	*$29.80*	-
Total Cost:	*$785.70*	*$682.95*	*$696.30*	*$120.19*

Figure 1

2) Intra-institution heterogeneity

- Compare different sources of “Same data”
  - Diagnoses
  - Labs
  - Meds

Heterogeneity of labs
Mayo

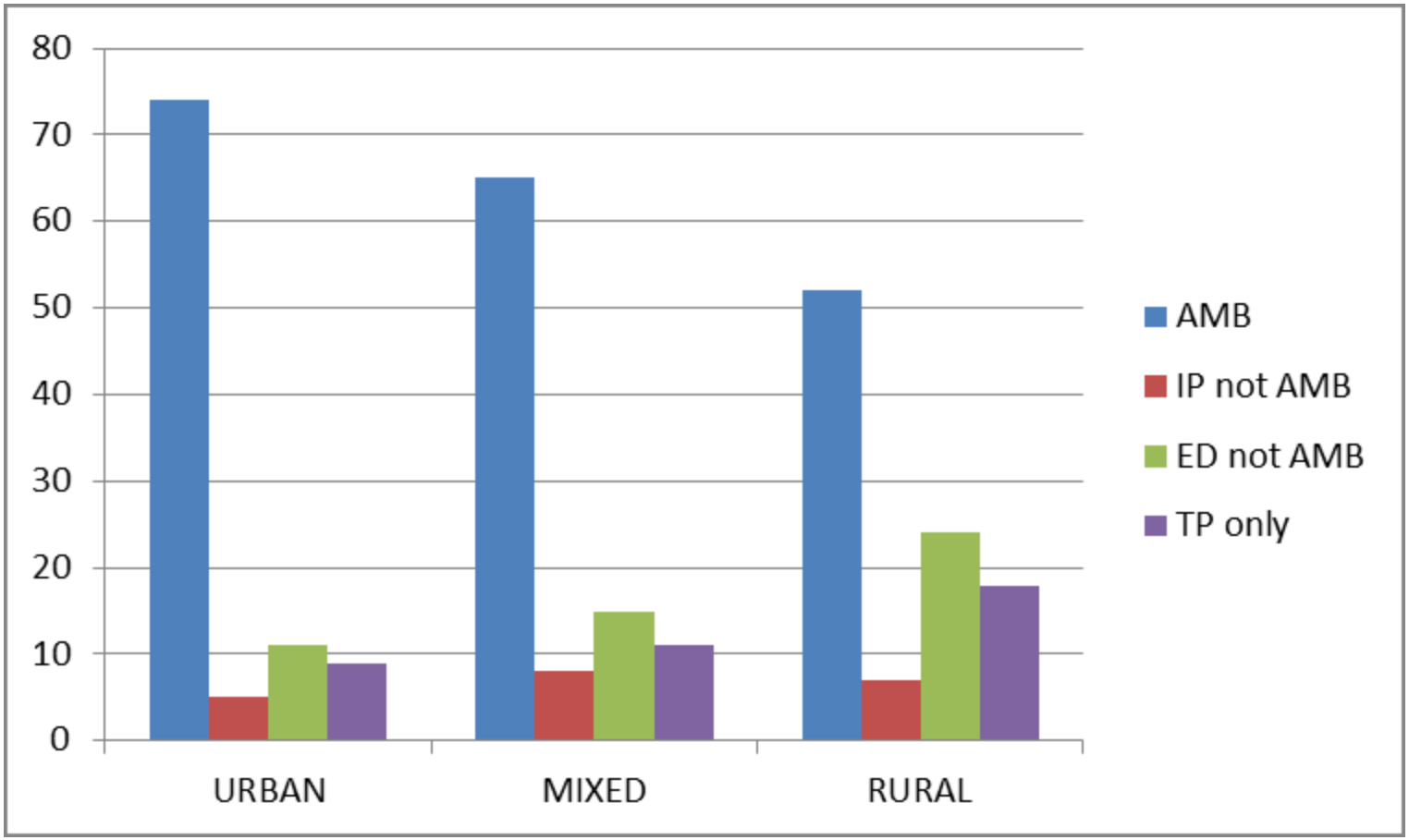
Glucose Labs: Mayo Laboratory Information Systems

"Box plots show significantly different distributions for lab test codes all identified as glucose. Only 1 code was clearly identified as fasting, while the others may have included, but not distinguished, fasting tests."

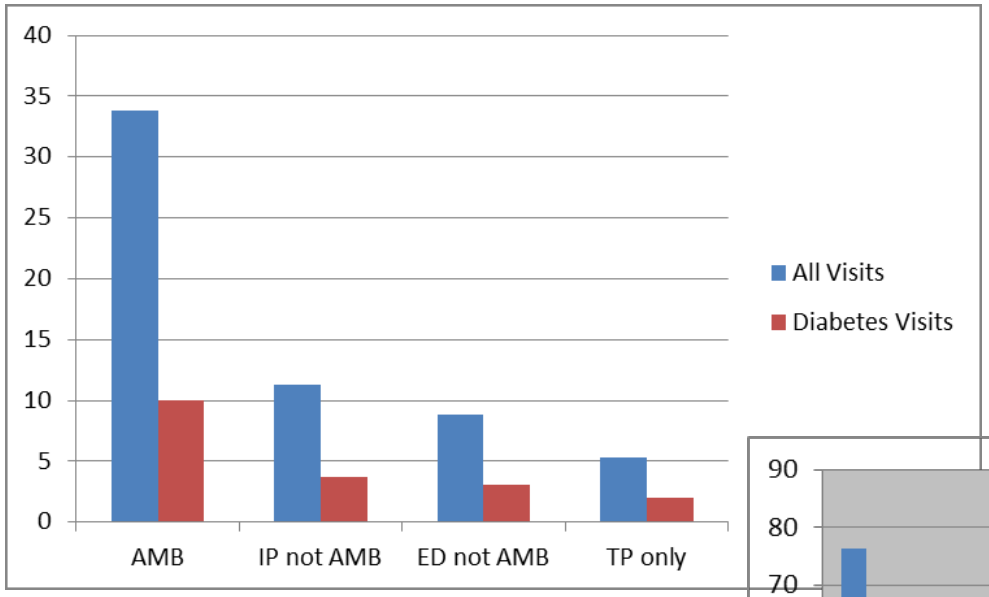
Intra-institution comparison of data sources Mayo Billing data vs. Problem List

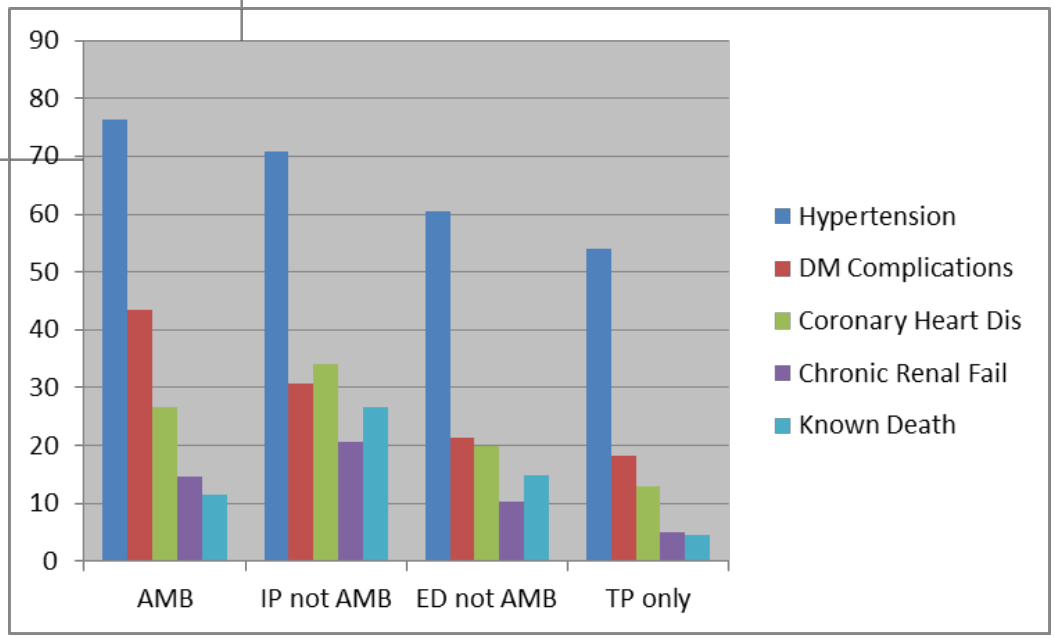
DM by Problem List	DM by Decision Support System		
	No	Yes	Total
No	2497	4981	7478
Yes	286	6355	6641
Total	2783	11336	14109

Intra-institution comparison of data sources IMtn regional differences in data



Intra-institution comparison of data sources IMtn Type of Encounter differences





Intra-institution comparison of data sources IMtn data recorded IP v Amb on same random adult cohort (n = 3191) with IP and Amb encounters 2007-2011

HIC3	FDB HIC3 Description	Freq IP data (HELP1 TRANORD)	Freq Amb Data (Med List)	Likelihood Noted In IP vs Amb Data
W7C	INFLUENZA VIRUS VACCINES	8	1.1	87.6
W7L	GRAM POSITIVE COCCI VACCINES	6.4	0.4	93.8
A4A	ANTIHYPERTENSIVES VASODILATORS	5.1	0.9	85.1
D6S	LAXATIVES AND CATHARTICS	65.5	13.2	83.2
H6J	ANTIEMETIC/ANTIVERTIGO AGENTS	51.5	16	76.3
C1D	POTASSIUM REPLACEMENT	39.4	10.3	79.3
C4G	INSULINS	19.7	5.5	78.2
Z2D	HISTAMINE H2-RECEPTOR INHIBITORS	19.5	4.4	81.4
A2A	ANTIARRHYTHMICS	10.9	2.4	82.2
Q5H	TOPICAL LOCAL ANESTHETICS	11.1	2.1	84.2
S2B	NSAIDS CYCLOOXYGENASE INHIBITOR - TYPE	62.1	32.8	65.4
W1W	CEPHALOSPORINS - 1ST GENERATION	43.1	16.3	72.5
P5A	GLUCOCORTICOIDS	34.7	17.2	66.8

3)Inter-institution heterogeneity of sources

Mayo/Intermountain comparisons

- Example of template for inter-institutional data source comparison

Type of Data	Source	Mappings	Recommend to use?	Outp/ Inp/ ED?	Provider specialty available?
Labs (Mayo)	LIS	LOINC	Yes	I/O/E	No
Labs (IM)	LIS+POC	LIS ->LOINC	Yes	I/O/E	Yes
Labs (IM)	POC->EHR	IM codes -> LOINC	Yes	O	Yes

Mayo/Intermountain comparison of BMI data availability

	IMtn-DM	IMtn-all	Mayo-DM
Target population	44723	> 1 million	11334
Sample size	2975	3107	
Average age	60	47	62
Male percent	51	42	53
Female percent	49	58	47
Num w/ wt&ht or BMI	2343	1723	9816
Prcnt w/ wt&ht or BMI	79	55	87
Average BMI median/patient	33.0	29.2	33.0
Median BMI median/patient	31.8	27.9	32.0

		IMtn-DM		*IMtn-all*		*Mayo-DM*	
		BMI recorded					
		YES	NO	YES	NO	YES	NO
		n = 2784	n = 191	n = 2426	n = 681	n = 9816	n = 1518
ICD-9-CM	CODE DESCRIPTION	FREQUENCY (% of patients having at least one ICD-9-CM code)					
250	Diabetes mellitus	100	100	11	4	100	100
272	Disorders of lipoid metabolism	71	45	29	7	78	47
401	Essential hypertension	72	49	29	10	73	53
719	Disorders of joint	22	12	20	12	32	21
278	Obesity and hyperalimentation	8	6	< 5	< 5	27	11

4) Machine Learning

- Use of Frequent Item Sets & Associative Classification methods
  - Discover most ‘interesting’ (mathematically) differences in two large data sets
  - Applied to measure differences in sources of data
  - Will apply to compare Mayo, Imtn data sets

- Susan Rea, InterMountain