Data Quality in Clinical Research

Presentation to Alzheimer's Disease Research Centers
September 10, 2010
San Francisco, Ca
Meredith Nahm, PhD, CCDM

Which of the following is NOT true?

1. A space-craft was destroyed in the atmosphere of Mars due to a data quality problem.
2. Popeye ate spinach because of the vegetable's high iron content.
3. The warmest year on record was misreported as 1998 because of a NASA Y2K problem in climate data.
4. The USS Vincennes accidently shot an Iranian civilian flight in 1998 killing 290 people due, in part, to a data error.
5. Datasets posted by the National Snow and Ice Data Center contained erroneous data for 16 years until discovery in 1996.
6. Kenneth Schustereit was turned down for job by Home Depot because a 30 year-old misdemeanor was reported as a Felony by ChoicePoint data brokers.

But there was a data error involved

A 1930's manuscript presented iron content of spinach in a way that the high numbers from dry spinach were from leafy spinach.

The misconception that Spinach is unusually high in iron continues today.

Referred to as **SPIDES**

Reprinted from: Sutton, M., SPINACH, IRON and POPEYE: Ironic lessons from biochemistry and history on the importance of healthy eating, healthy skepticism and adequate citation. Internet Journal of Criminology, 2010.

Topics

- Defining data quality
- Data sources
- Impact of data processing on accuracy
- Errors & counting
- Data accuracy in
  - Medical record abstraction
  - Patient reported data / questionnaires
- Take aways

Defining Data & Information Quality

- **Accuracy**
- Currency
- Completeness
- Consistency
- Timeliness
- Relevance
- Granularity
- Specificity-definition
- Precision
- Attributability

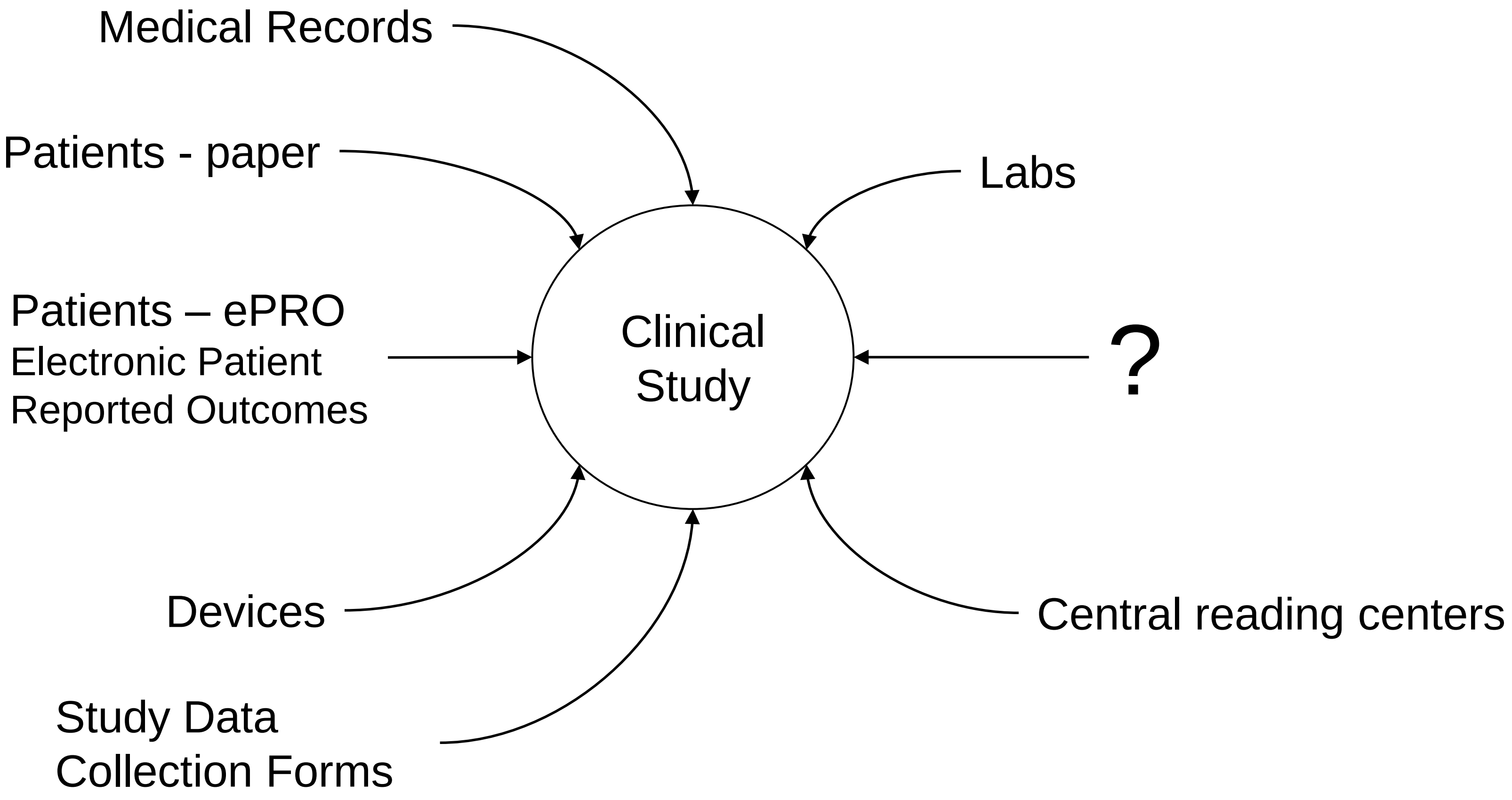
This presentation concentrates on accuracy.

Consider These Things

1. Data **source**
   – Each data source has it own quality issues
2. Data **processing**
   – Causes and corrects data errors
3. Interaction between the two
   – Data origin influences how data are processed

We will look at how these impact data quality.

Where Do Data Come From?



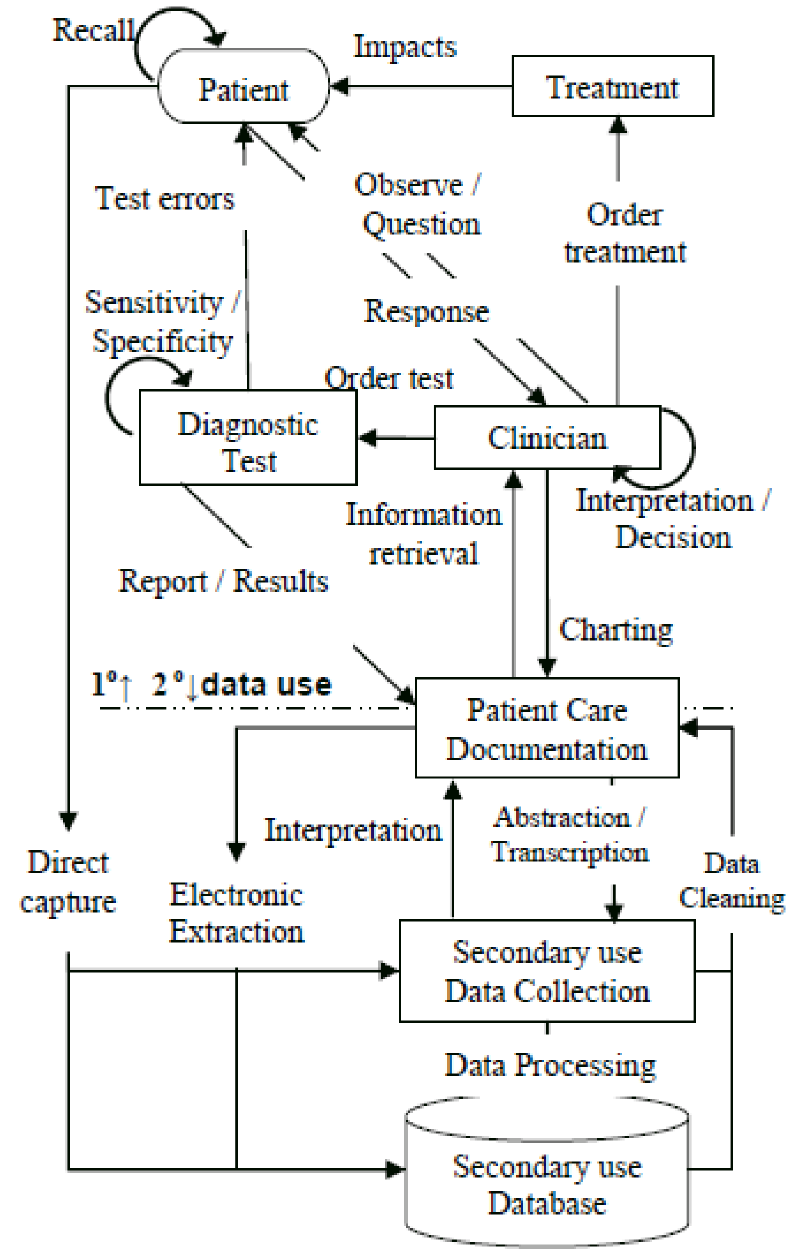
Information flow in healthcare and from healthcare to secondary uses gets complicated fast !

Nahm, M., Johnson, C., Johnson, T., Zhang, J., What Happens to Data Before Secondary Use? AMIA 2010.

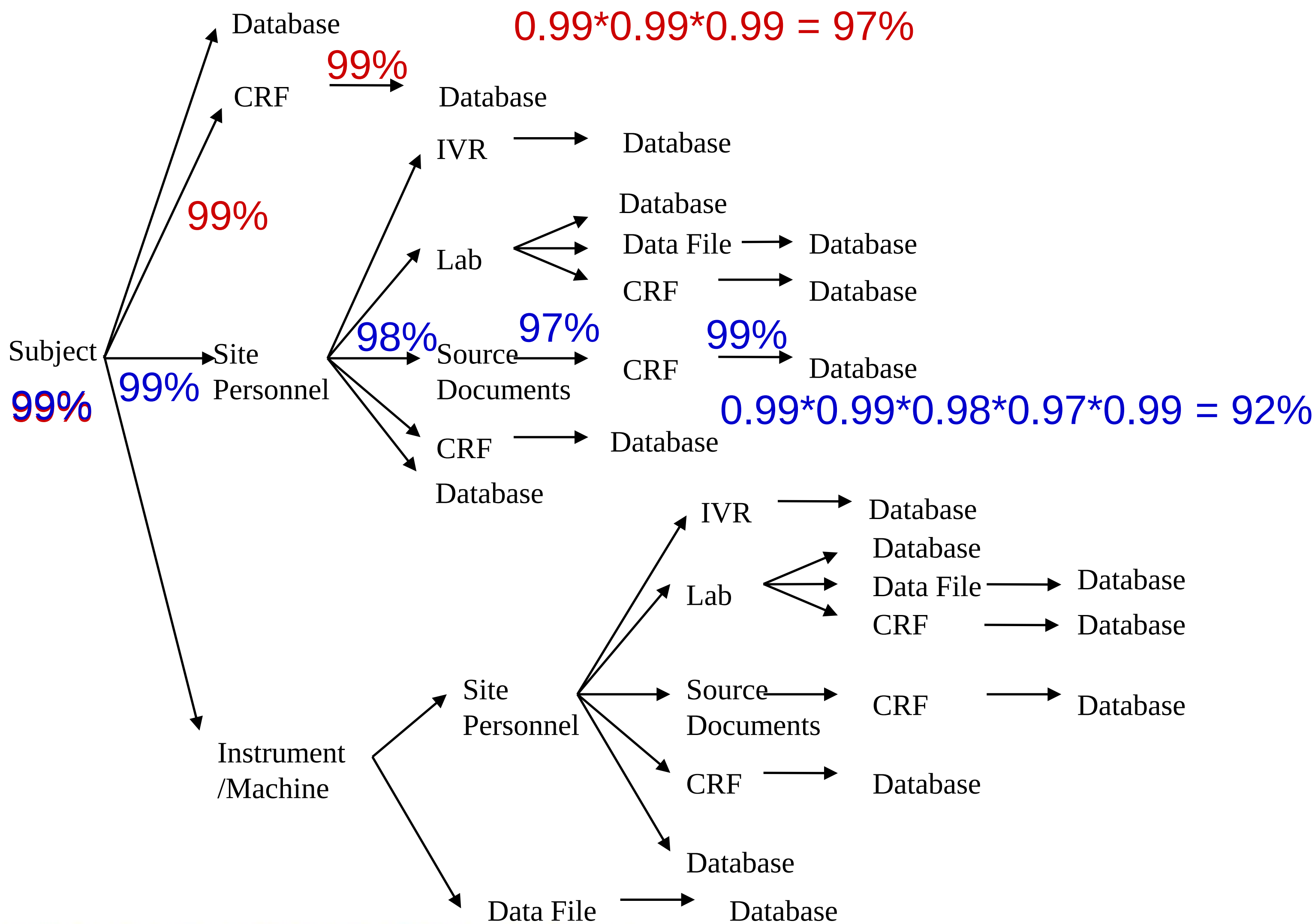
Synthesis of four published models.

References

1. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc*. Sep-Oct 1997;4(5):342-355.
2. Wyatt J. Acquisition and use of clinical data for audit and research. *J Eval Clin Pract*. Sep 1995;1(1):15-27.
3. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: Where are the methods? *Ann Emerg Med*. Mar 1996;27(3):305-308.
4. Nagurney JT, Brown DF, Sane S, Weiner JB, Wang AC, Chang Y. The accuracy and completeness of data collected by prospective and retrospective methods. *Acad Emerg Med*. Sep 2005;12(9):884-895.

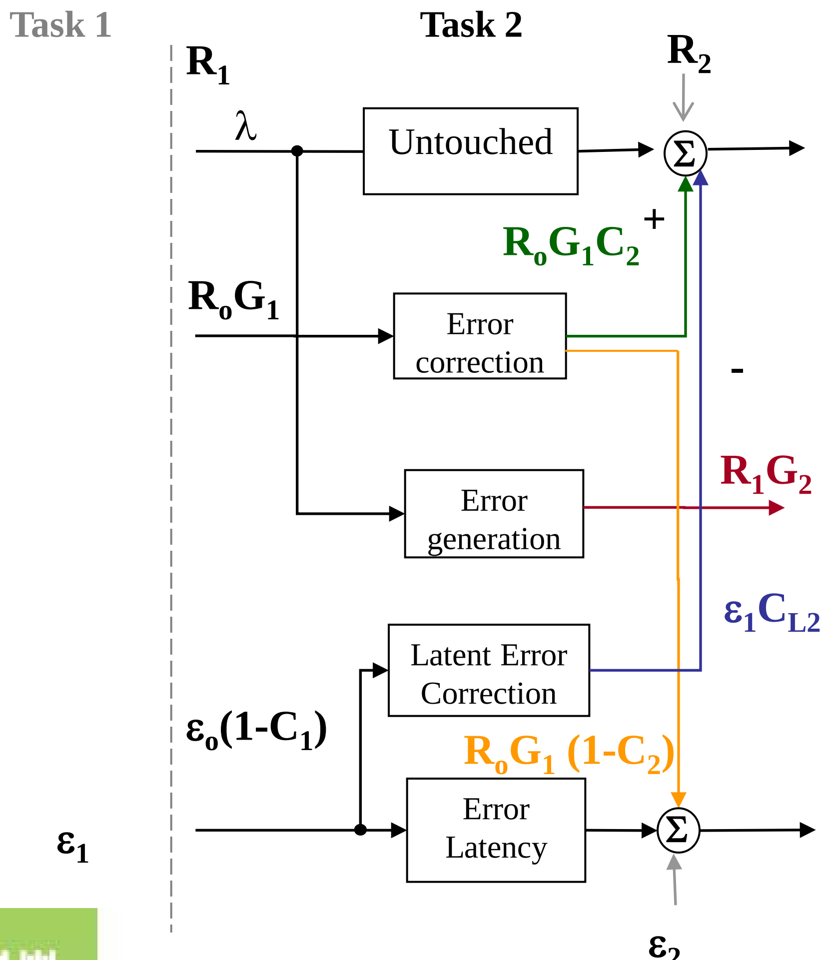


Data Paths in Clinical Research



Duke Center for Health Informatics

Taking Into Account Correction Rates



Task 3

R_o and ε_o are the input or initial values.
Each task may touch all or a fraction, λ, of the data.
R_i represents the number of accurate data values outgoing from the i^{th} process step.
ε_i, is the number of values in error outgoing from the i^{th} process step.
G and C are the error generation and correction rates

$$R_i = R_{i-1} - R_{i-1}G_i + R_{i-2}G_{i-1}C_i + \varepsilon_{i-1}C_{iL}$$

$$\varepsilon_i = \varepsilon_{i-1} - \varepsilon_{i-1}C_{iL} + R_{i-2}G_{i-1}\left(1 - C_i\right)$$

Error Latency

Duke Center for Health Informatics

Example Application

Input
R_o = 1000 fields
ε_o = 100 errors
0.09 or 9% error rate

Three step process common in clinical research including 1) chart review (medical record abstraction), 2) data entry and 3) data cleaning. Input data stream comes from medical records with 1000 accurate fields and 100 fields in error.

Task	Error Generation Rate (G_i)	Error Correction Rate (C_i)	Latent Error Correction Rate (C_{iL})	Outgoing Number of Accurate Fields (R_i)	Outgoing Error Number* (ε_i)
Chart review	0.03	0.01	--	971	99
Data entry	0.0025	0	0	969	129
Cleaning	0.0025	0.01	0.01	968	132

*delayed accumulation due to modeling error generation in one step as input to error correction of next step.
Shaded areas are input to the model.

Notice that for G and C indicative of our industry, as data processing steps are added:
- Outgoing number of accurate fields decreases
- Outgoing error number increases

Ideal = low error generation rates and high correction rates

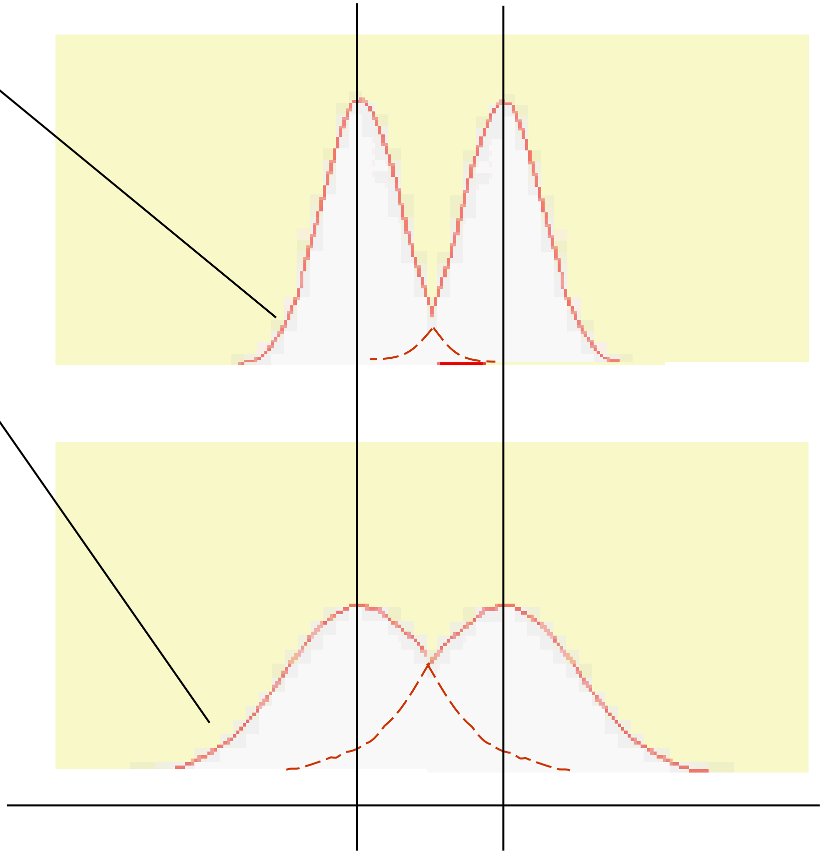
Options

- Measure the rates for your processes
- Use literature values
- Guess & calculate upper and lower bounds

Why are Errors Undesirable?

Fewer errors, less variability (narrower distribution)

More errors, More variability (wider distribution)

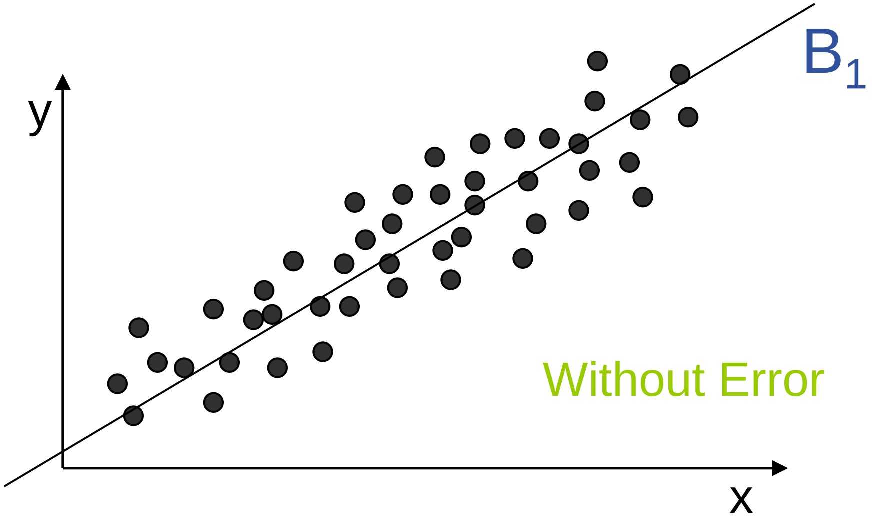


- Errors can introduce variability into the analysis
- Variability makes it more difficult to see a "treatment effect", if there is one

i.e., errors can bias toward the null …

Duke Center for Health Informatics

Regression



- $B_1^* < B_1$
- B_1^* biased toward the null
- Standard error of B_1^* is larger
- Therefore, p-values are larger
- Power is less, alternatively
- Sample size required to maintain the same power ↑

Important Things to Know

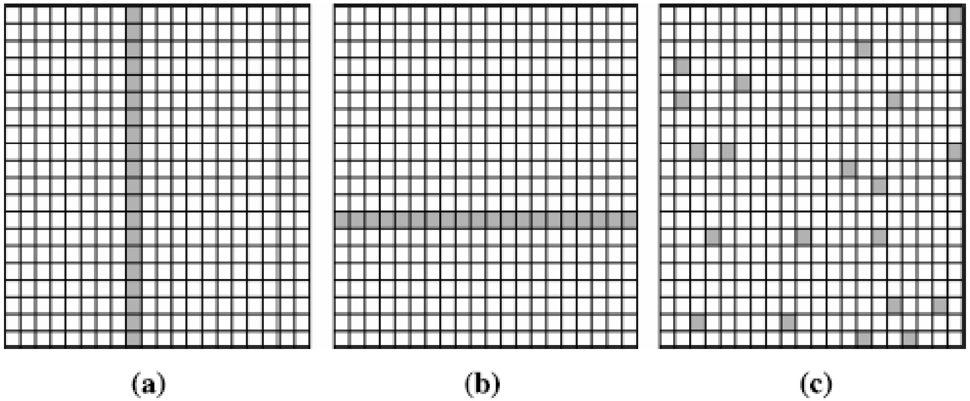


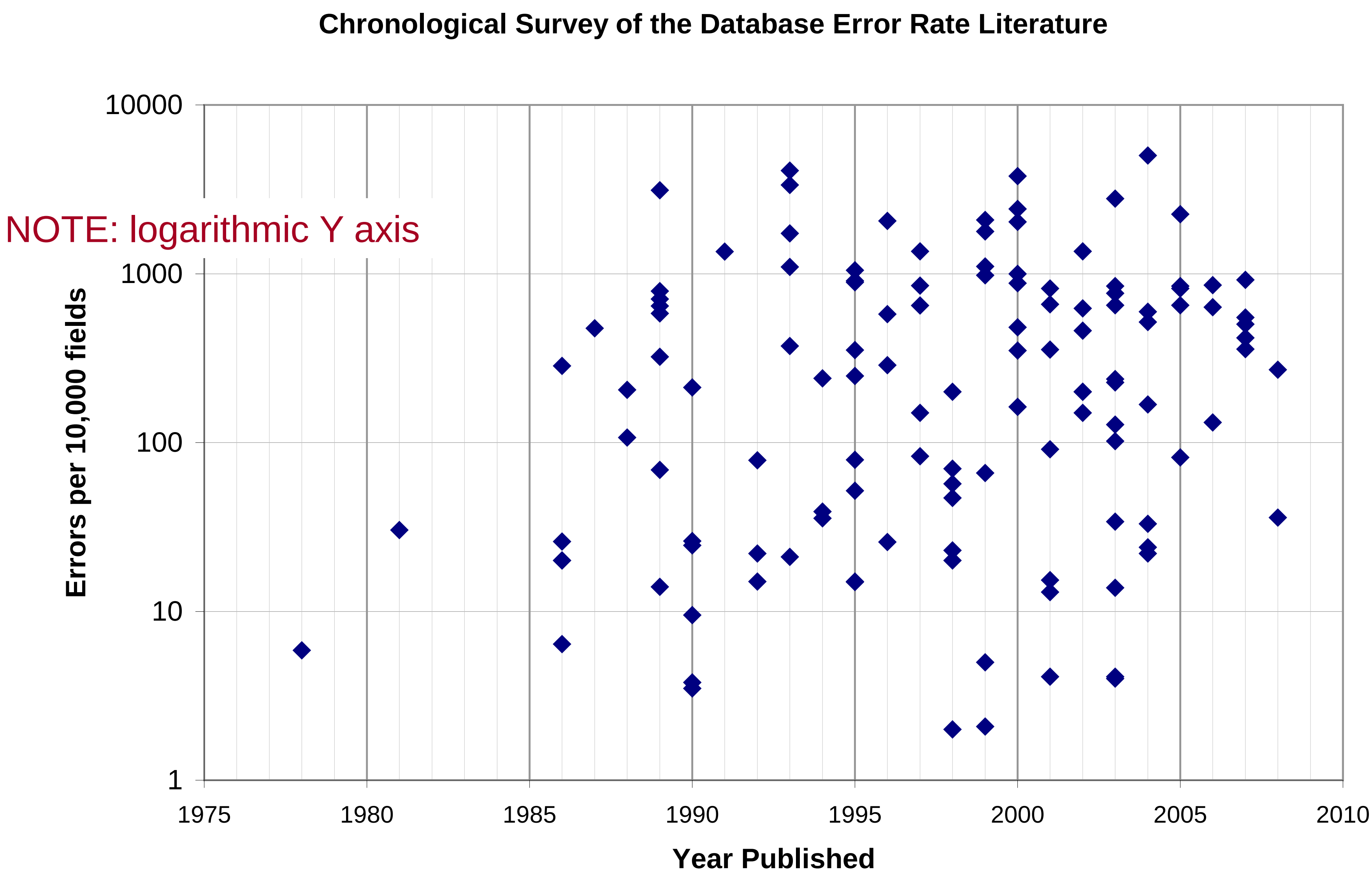
Fig. 1. Distribution of errors.

1. Error rate
2. Distribution of errors (a, b, or c)
3. Distribution of error values with respect to other values (bias, in-range, outliers)

Diagram from Fisher, C.W., Lauria, E.J. M., Matheus, C.C., An Accuracy Metric: Percentages, Randomness, and Probabilities. ACM Journal of Data and Information Quality, Vol. 1, No. 3, 2009.

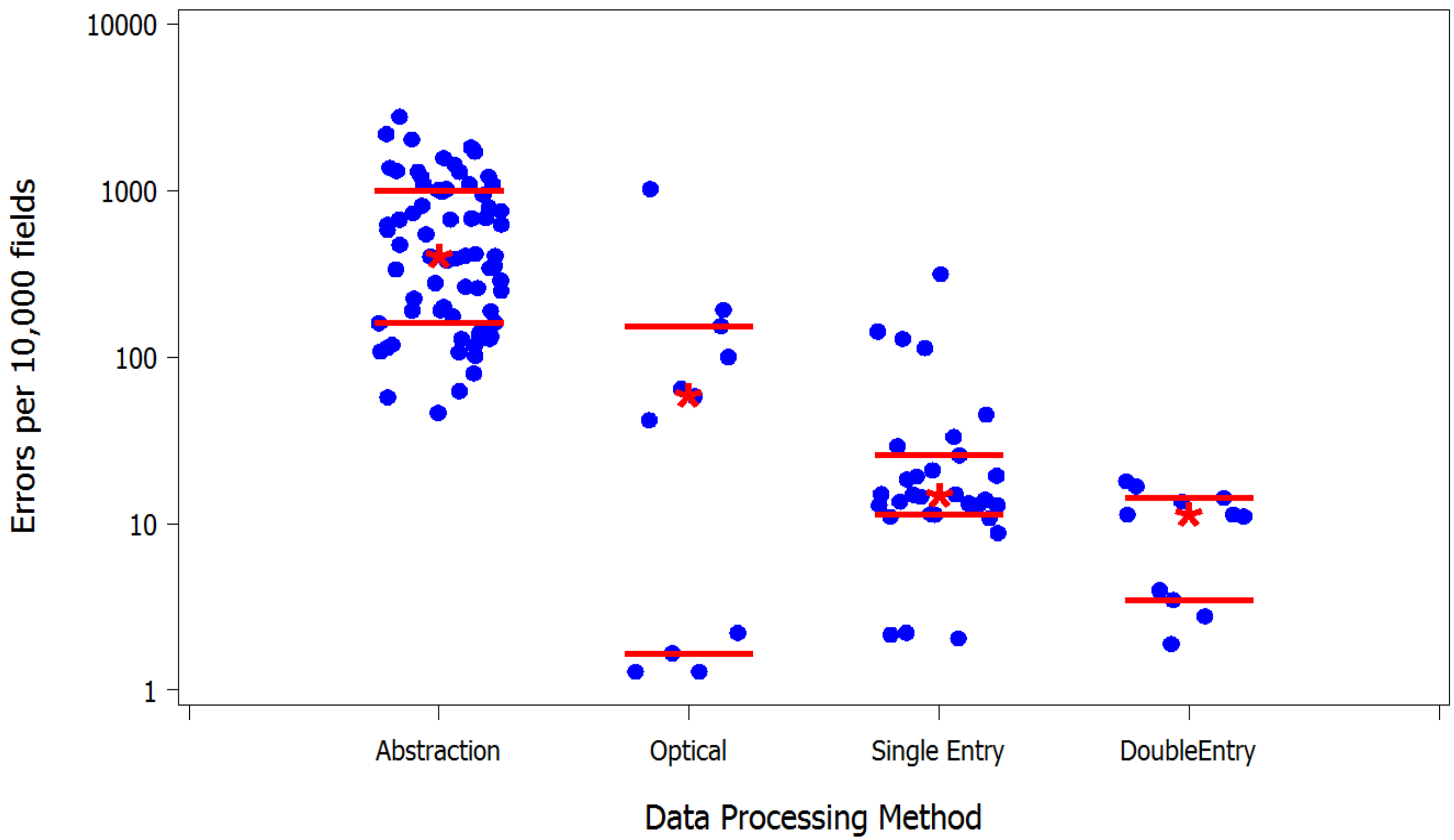
What do we know from the literature that might help us ?

Results from two systematic literature reviews with pooled analysis.



Duke Center for Health Informatics

Figure 1: Data Error Rates for Different Processing Methods



Duke Center for Health Informatics

What We Know So Far About Data Processing

- Different data collection and processing methods are associated with:
  - different median error rates
  - different dispersion
- Ranges overlap → other factors involved
  - Single entry can perform as well as DDE
  - Optical methods can perform as well as DDE
- Medical Record Abstraction associated with highest & most disperse error rate

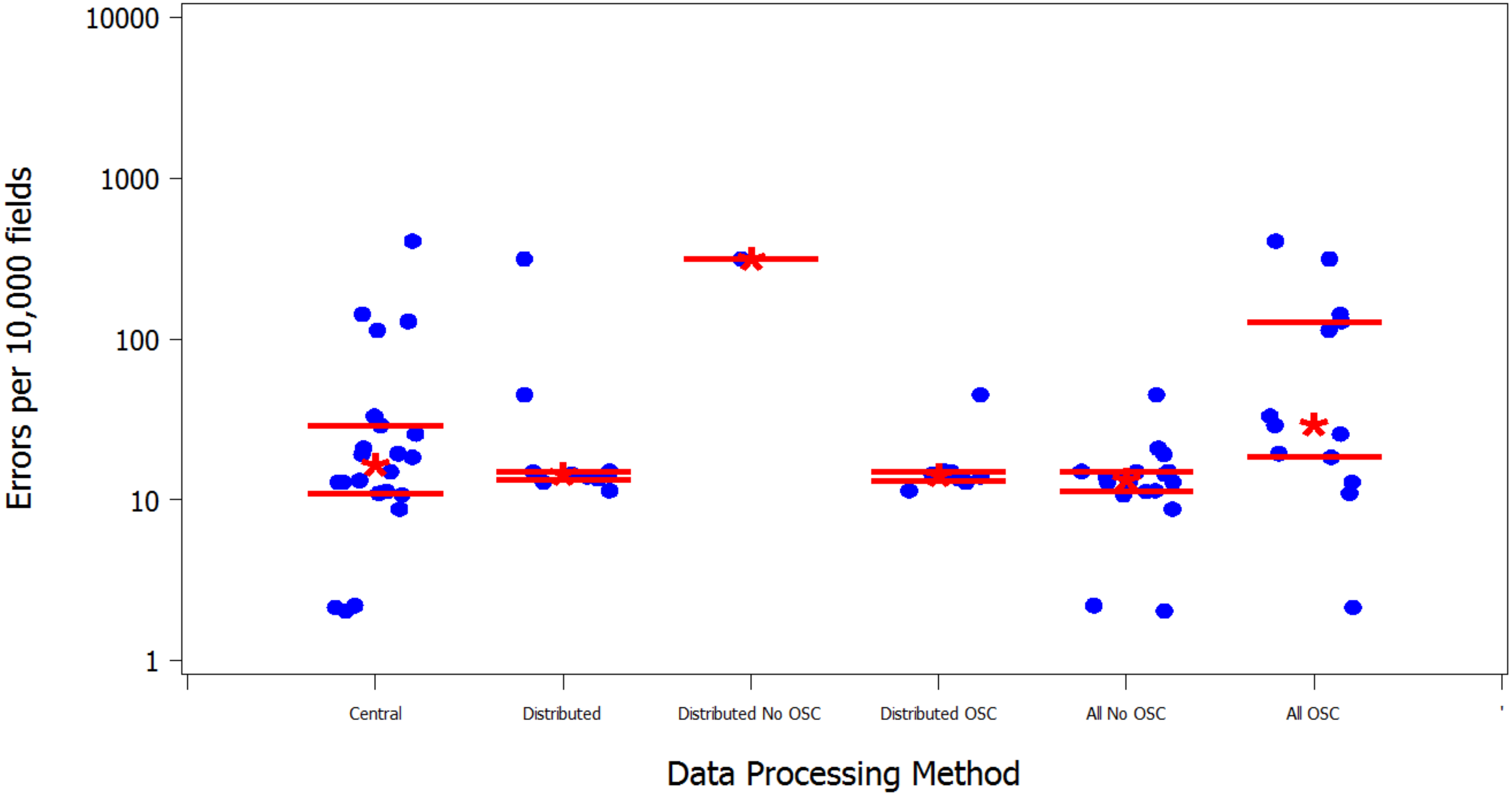
Data Collection & Processing

- **Add to** source errors
- **Is added to** by further data processing

Don't Touch My Data !

Duke Center for Health Informatics

Figure 2: Error Rates from Central and Distributed Data Processing in Presence of On-Screen Checks

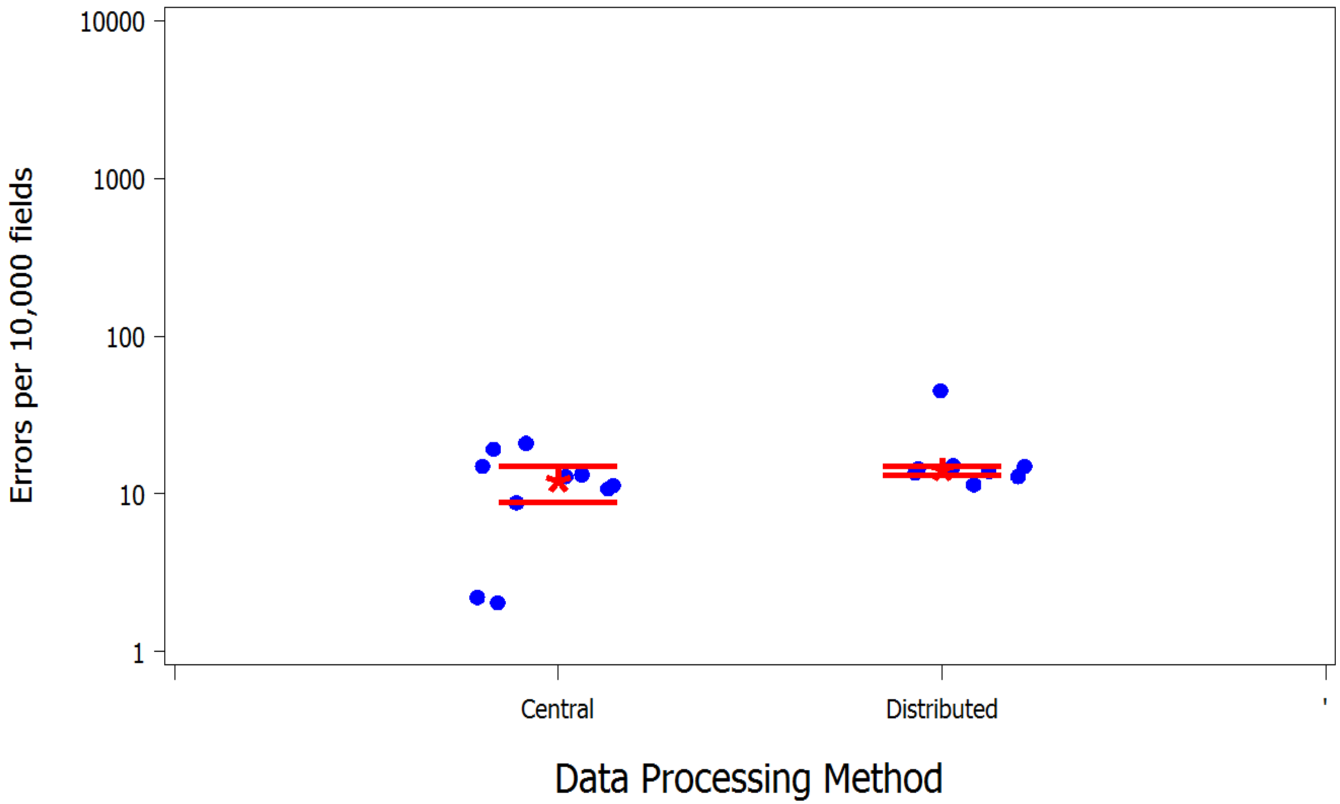


Duke Center for Health Informatics

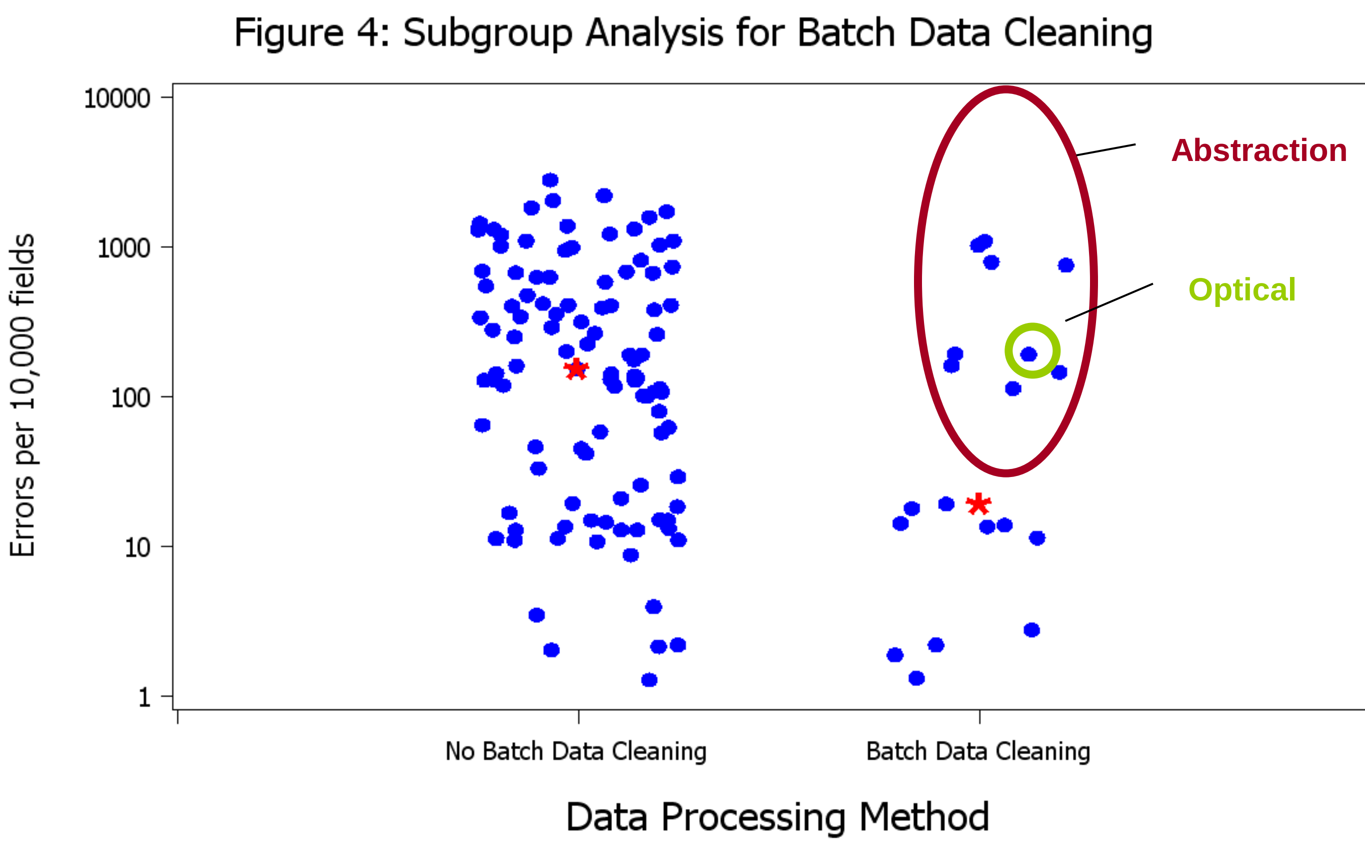
Error Rates from Central and Distributed Data Processing In the Presence of On-screen Checks

EDC Model:

i.e. single entry with on-screen checks



Subgroup Analysis for Batch Data Cleaning



Duke Center for Health Informatics

Data Cleaning

- Location of data processing has little effect
- On-screen checks in single entry effective → bodes well for web-based EDC
- In presence of OSC, distributed data processing is associated with error rates comparable or better than central processing
- Batch cleaning, while effective
  - Is expensive 1-10-100 rule, $50-100 per query
  - Can not reach errors in source or abstraction errors

Auditing and Audit Timing

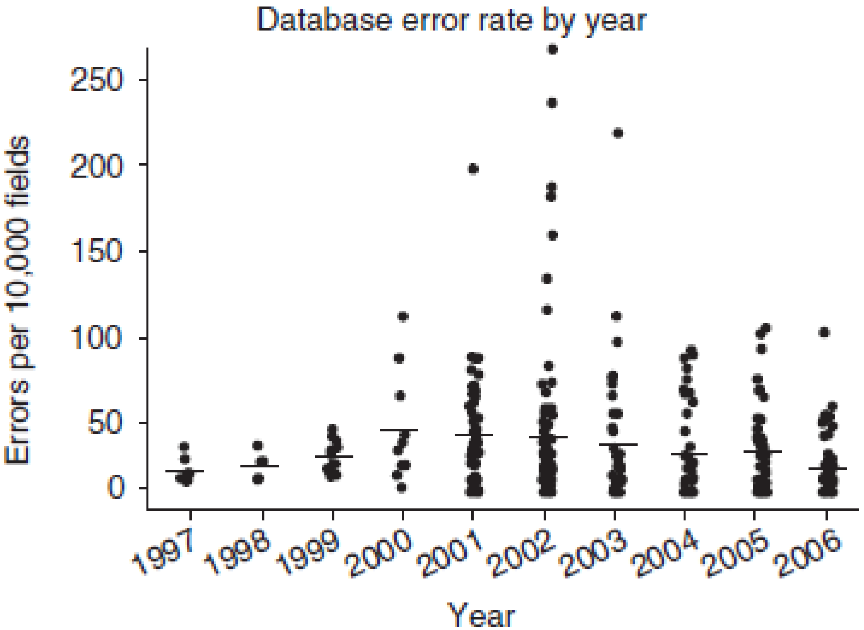


Figure 2 Database error rates by year, 1997–2006. From 1997 to 2001, variability and the central tendency of the error rates for all database-lock audits performed prior to start of SPC audits trend upward. After SPC implementation in third quarter 2001, data variability declined and quality increased

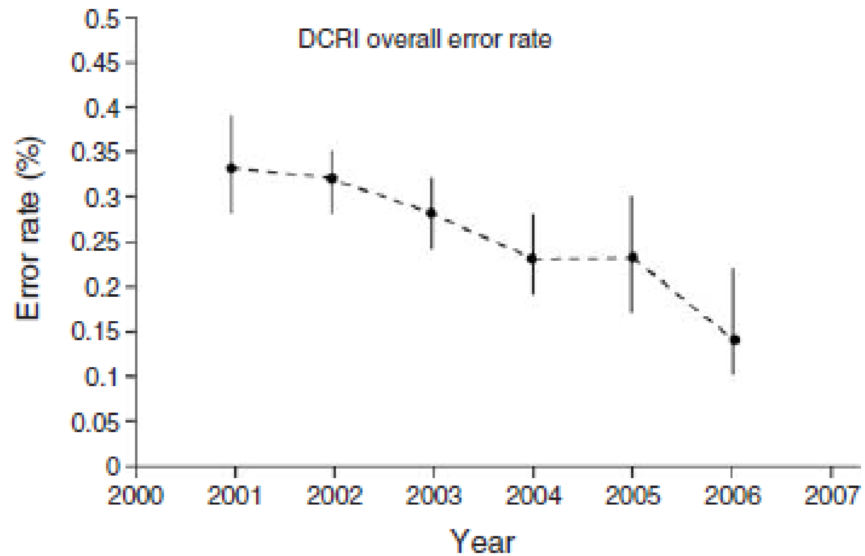


Figure 4 DCRI overall error rate for SPC audits, 2001–2006. Data show a downward trend in error rates following the initiation of SPC audits (and therefore an increase in data quality)

A few small audits are better than one pre-database lock audit.

Rostami,R., Nahm M., Pieper, C.F., What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997–2006. Clinical Trials 2009; 6: 141–150.

We Can’t Count …

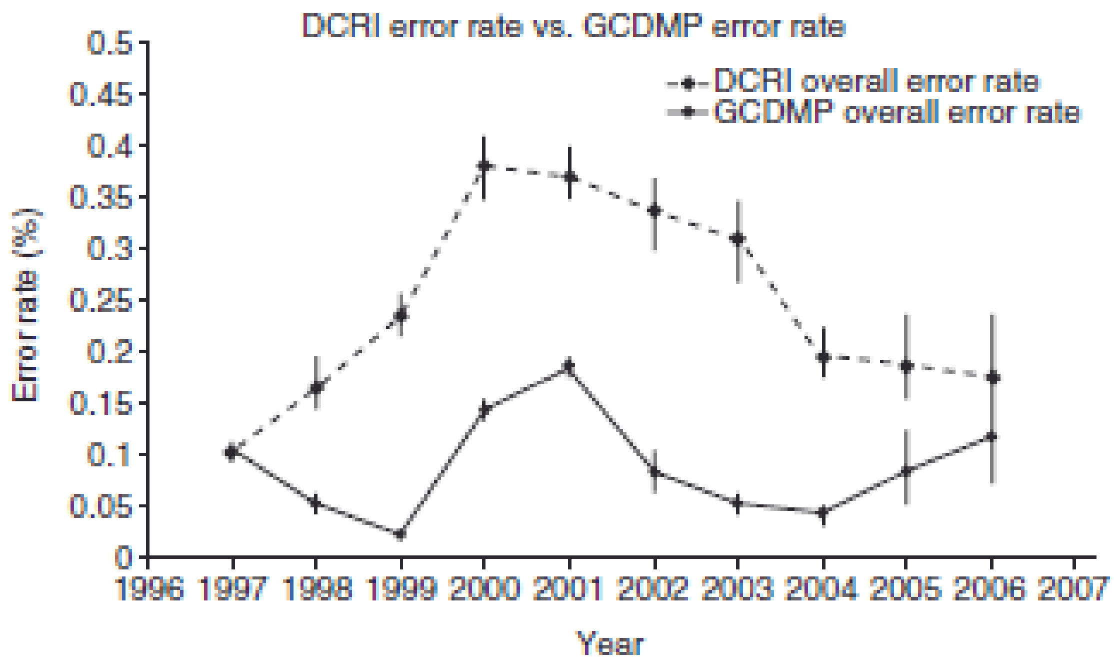
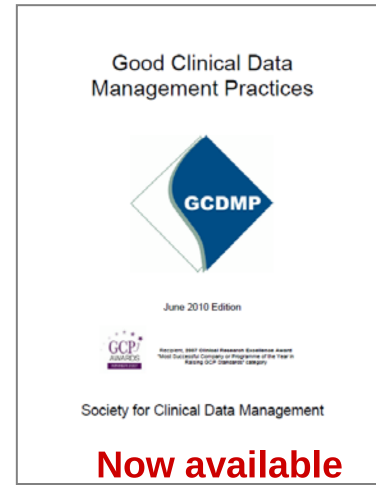


Figure 6 DCRI error rates compared with GCDMP error rates, 1997–2006. GCDMP methods consistently yield lower error rates. Differences range from a factor of 1.5 to a factor of 9.5; the variance is a function of degree of sparseness of data on CRFs



Now available in Japanese

We Really Can't Count …

Table 2 Reliabilities for various distributions and proportions for analytical approach shown in Figure 7

Error/10,000	Error rate	No range check—uniform (0,999)		Range check—uniform (0,300)	
		Reliability	% sample increase	Reliability	% sample increase
5	0.05%	0.844	18%	0.984	2
10	0.10%	0.730	37%	0.968	3
50	0.50%	0.351	185%	0.857	17
100	1.00%	0.213	370%	0.750	33
500	5.00%	0.051	1848%	0.375	167

Bottom line: Data Accuracy matters !

Now, Who Can't Count ?

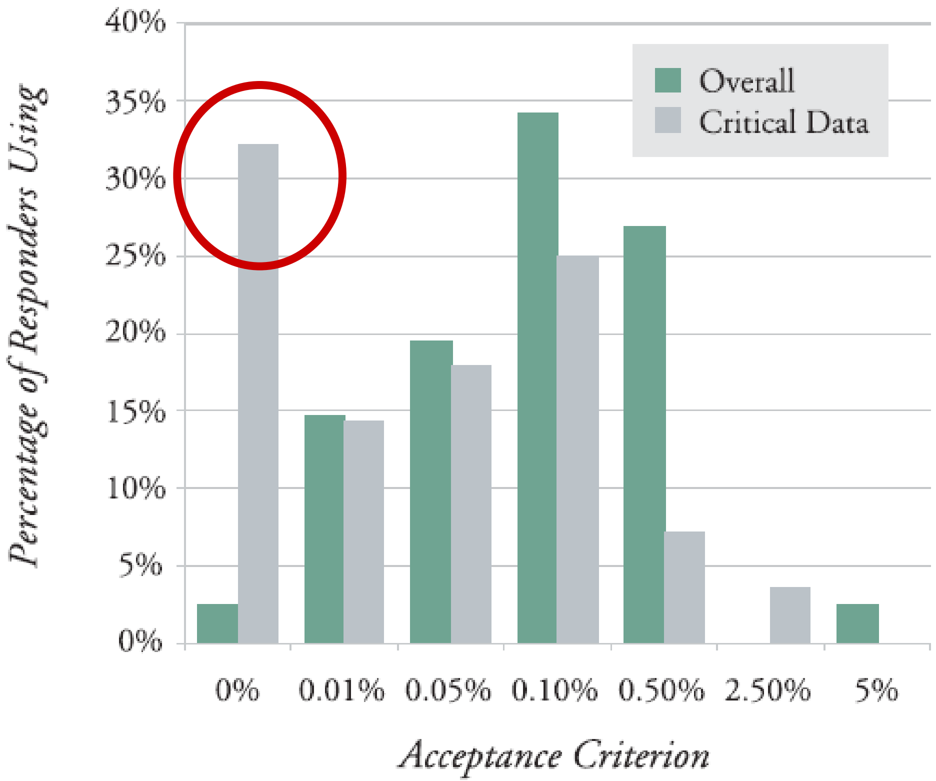


Figure 1: Acceptance Criterion for Critical Data and Whole Database

The most popular acceptance criterion from the SCDM 2004 Data Quality survey:

<u>Overall database error rate</u> - 0.10% and 0.50%, or 10 and 50 errors per 10,000 fields

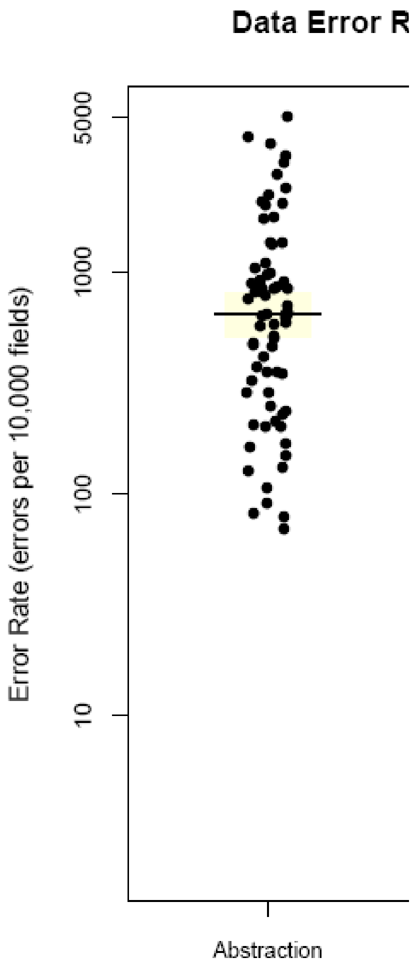
<u>Error rate in critical fields</u> – **0%** and 0.10% or 0 and 10 errors per 10,000 fields.

At a recent industry conference, I asked a few hundred industry data managers, "Who audits source-to-CRF?". Two people raised their hands …

Duke Center for Health Informatics

(Nahm, *et al.* Databasics, Summer 2004.)

Did I Mention Medical Record Abstraction Error Rate?



Duke Center for Health Informatics

Review of the MRA Literature

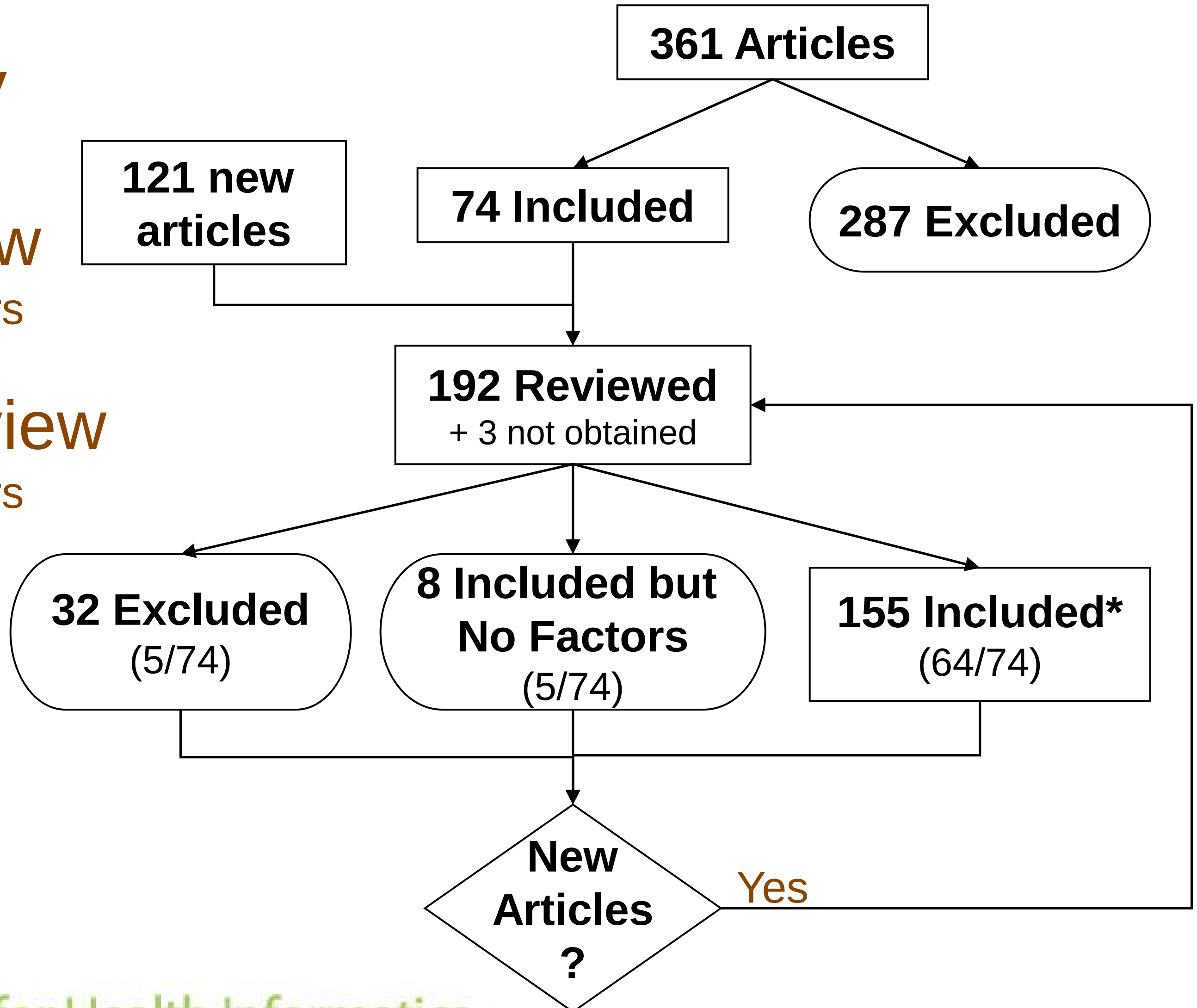
PubMed Query

Abstract Review

2 independent reviewers

Full Article Review

2 independent reviewers



Duke Center for Health Informatics

*Six RAND Reports lumped into one

Cause-Error-Mitigator Model

Candidate causes

Cause(s)
One of which may be root. Causes may be of human or non-human origin

Aggravators
↑ # or impact of errors

Mitigators
↓ # or impact of errors

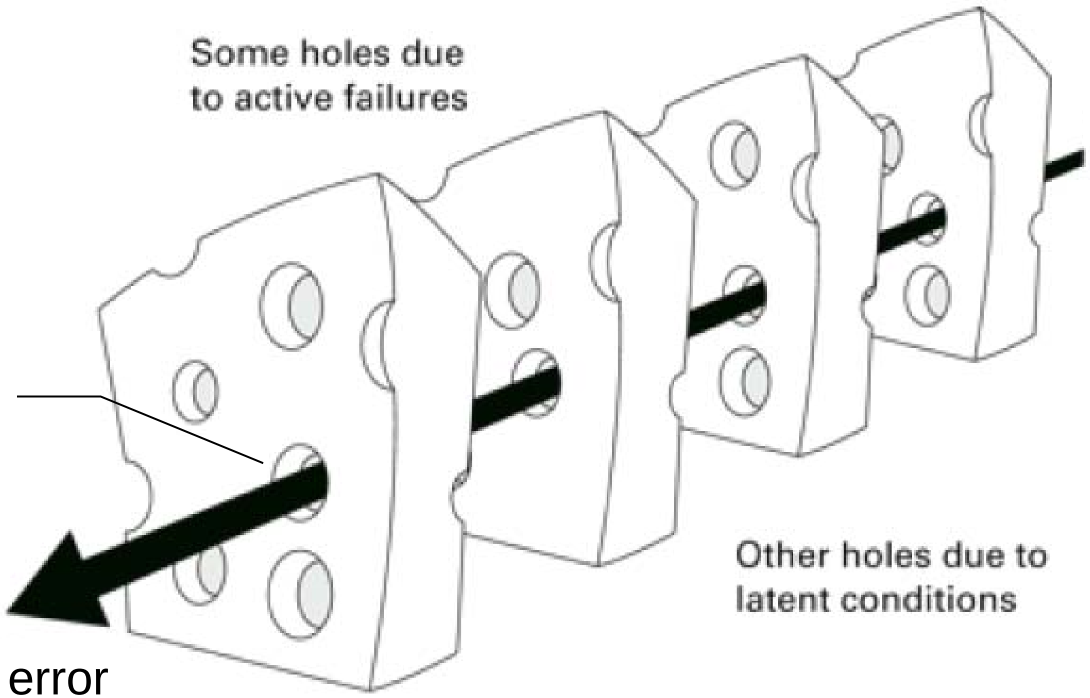
Error(s)
May be observable or not observable,
May be controllable or not controllable,
May be observable but not controllable,
May be controllable but not observable, …
May be immediately obvious or latent

Impact of this model on this research:
Need to retain both valence of factor statement as well as direction of impact, *i.e.,* increases or decreases accuracy.

valence	direction of impact	
	+ ↑	+ ↓
	- ↑	- ↓

Duke Center for Health Informatics

Swiss Cheese model[2]

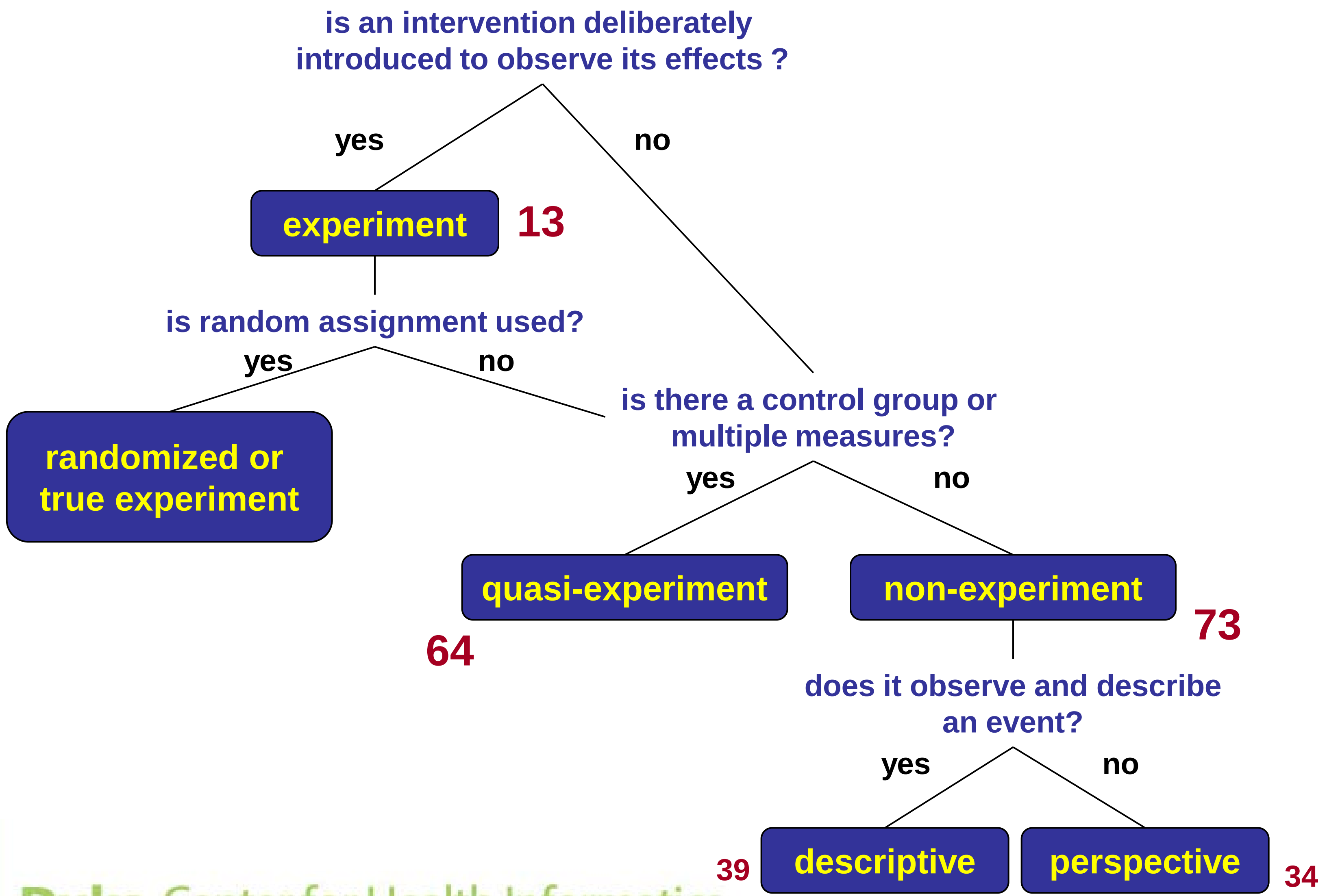


“Sharp end error”

Accident / error

2. Reason J: Human error: models and management. *BMJ* 2000 , 320:768-70.

Classification of Literature (n=150)



Duke Center for Health Informatics

Adapted from: William M.K. Trochim, accessed March 28, 2010. http://socialresearchmethods.net/kb/destypes.php

Literature Variables for
Experimental and Quasi-experimental work

- Independent : Data collection method
  - Abstraction *vs.* extraction
  - Abstraction *vs.* reabstraction
  - Abstraction *vs.* independent database
- Dependent : measure of error or discrepancy
  - Percent agreement
  - Kappa, Kappa variations
  - Sensitivity, specificity
  - ICC

Choice is based on your philosophy, *i.e.,* is there a gold standard or knowable truth to which you compare …

Literature Top 12 (overall)

292 unique factors,
2385 total mentions,
166 with >3 mentions

Factor	Mentions
Training abstractors	72
Standard data collection forms	57
Missing information in the medical record	55
Measuring inter- or intra abstractor reliability	51
Conventions for data elements	43
Reabstraction of data	42
Standard abstraction process	39
Variability in clinical documentation practices between clinicians and organizations	37
Definition of data elements to be abstracted	37
Independent data sources	36
Error in the medical record	33
Type of Data	33

Content Validity Assessed by Delphi Process

- Recruited at national conferences
  - 20 Clinical Research
  - 18 Registry
- Eligibility criteria
  - Three or more years of abstraction experience as reported by the participant
  - Abstraction experience in either a clinical research or registry / quality improvement setting
  - Able and willing to give informed consent

Duke Center for Health Informatics

Delphi Results: Top 12

1. Illegible information in the medical record
2. Training abstractors
3. Abstractor credentials*
4. Missing information in the medical record
5. Definitions for each data element to be abstracted
6. Access to charts*
7. Interruptions*
8. Conventions or guidelines for data elements
9. Experience in the clinical area
10. Limited time*
11. Conflicting information in the chart
12. Variability in clinical documentation practices between clinicians and facilities

* Higher relative frequency of mention in the Delphi

Factors identified in Delphi that were NOT Found in the Literature

Factor	Mentions
Interruptions	6
Complexity of the study or project	3
Supportive collegial relationships with physicians, nurses, and medical records colleagues	3
Abstractor (human) error	3

Delphi Refutes Literature

ratings of neutral (3.0) or less on a 5 point scale

Factor	CR	R / QI	Overall
RN credential	3.2	2.2	2.8
Blinding abstractors to study aims	2.5	1.9	2.2
Centralized abstraction	3.2	2.7	3.0
High study / project complexity*	3.8	2.5	3.3
Thick medical records	3.2	2.8	3.1
Patients cared for by multiple providers / facilities	3.6	2.8	3.4
Presence of multiple diagnoses / procedures	2.8	2.6	2.7

*Delphi factor NOT a Literature Factor

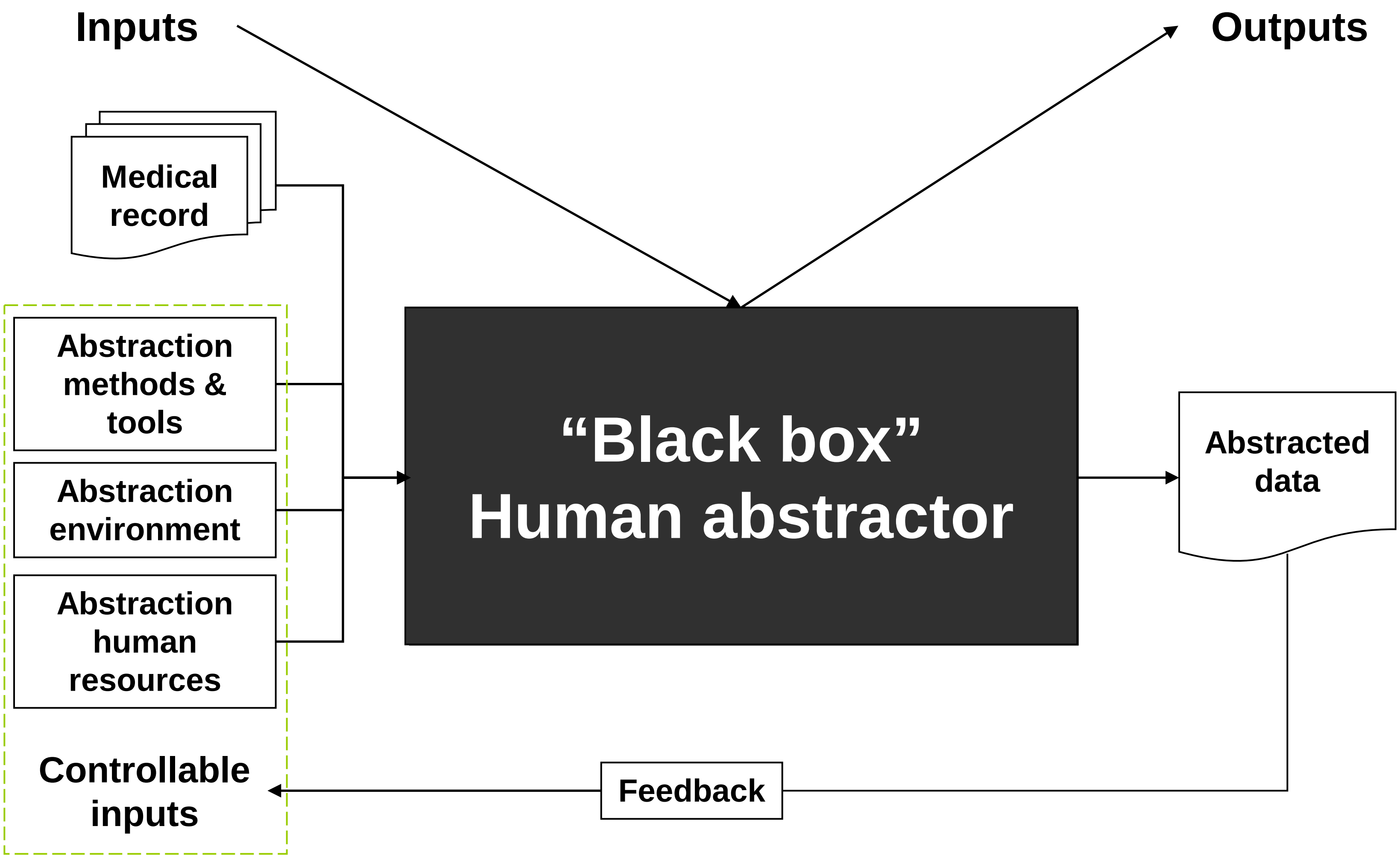
Conclusions

- Consistency between the two Delphis:
  - factors impacting accuracy are similar in Clinical Trials and Registries
- 296 factors (Literature + Delphi)
  - data accuracy in medical record abstraction is a complex, many-faceted problem
- From the high level (73%) of agreement between expert abstractors and the literature
  - The factors are real

So what ?

- MRA is largely an information loss and degradation problem
- The number of factors large (296)
- We have indication of which are most important
- Practitioner approach: heuristic based on the factors AND early assessment and monitoring of error rate

Medical Record Abstraction System



Heuristic & Monitoring

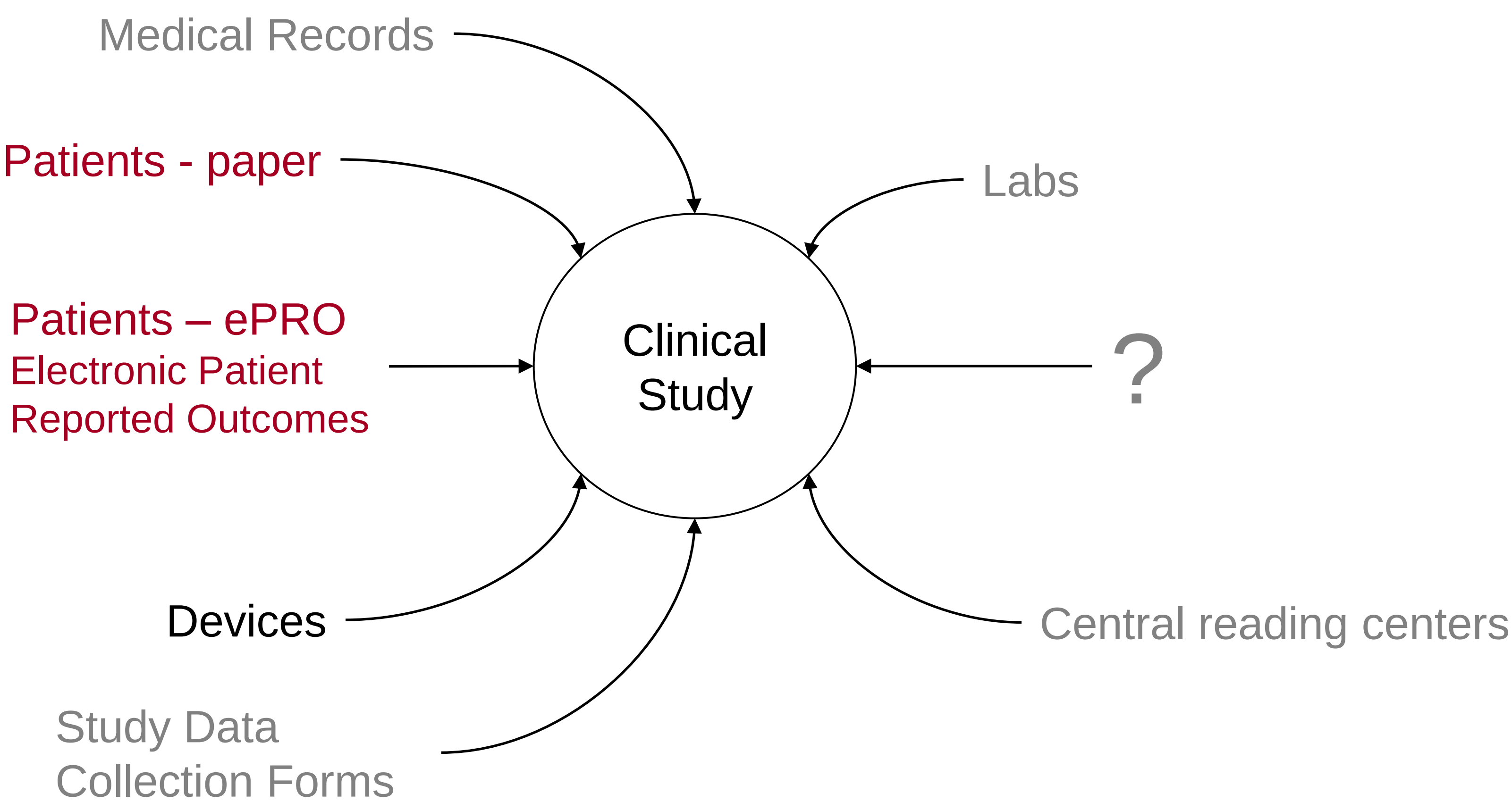
Assure

- Source
  - Potential errors/missing data based on types of data to be collected
  - Use original recording
  - Necessitates knowledge of clinical area
- Atomic data elements
  - Objective, completely defined, including source/s
- Standard data collection form
  - Representation & cognitive support
- Abstractor training with practice & feedback

Make sure

- Redundancy
  - Independent data sources, re-abstraction
- Issue identification & resolution process
- Pilot test the abstraction process
- Periodic monitoring through re-abstraction, IRR, RCA with feedback

What About Other Data Sources



Duke Center for Health Informatics

Data From Patients
Factors that Have Been Studied

- Reliability and Validity
- Interviewer versus Self-administered
- Paper *versus* electronic
- Mail *versus* internet *versus* telephone
- Question format
- Questionnaire length
- Recall bias
- Use of Proxies
- Paper versus electronic diaries
- Type of information - social stigma, major life events
- Fraud

Literature covers patient care, clinical research, psychometrics, marketing research, …

Paper *versus* Electronic

"Numerous anecdotes describe subjects completing a week's or a month's worth of diaries in the parking lot before delivering them to sites."[1]

- *Study 1: Asthma* electronic monitor data indicated 52% of measurements missing, manual diary card data indicated only 15% missing [1]
- Study 2: subjects invented 425 (22%) of values written in the diary card [1]
- Study 4: 67% of the subjects overestimated their compliance, and 30% of the diary entries were in error [1]
- Study 3: over 90% of subjects exaggerated use of inhaled corticosteroids [1]

1. Data taken from review article by Raymond S.A. and Ross R.N., Electronic Subject Diaries in Clinical Trials. Applied Clinical Trials, March 2000.

Type of Information

- Secondary cardiac events reported accurately, “Have you had a Heart Attack in the last 6 months?”
- Sensitive information (stigma or illegal activity)
  - General feeling not as accurate
  - NIDA CTN uses computer rather than interviewer

Data Processing

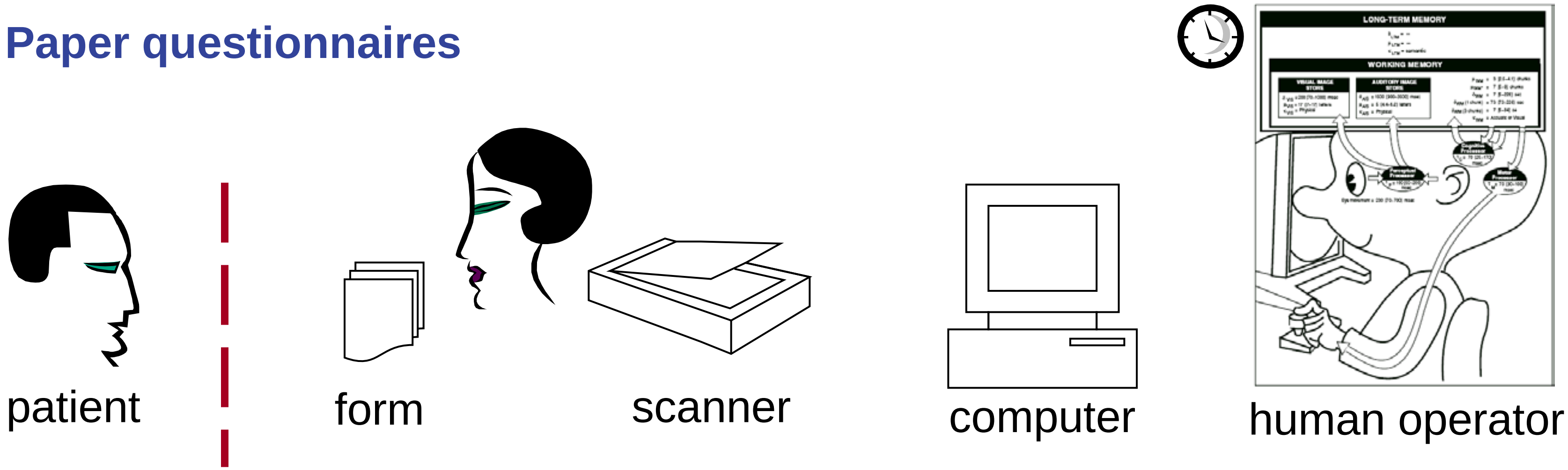
- NSABP (National Surgical Adjuvant Breast and Bowel Project) Breast Cancer Prevention Trial: questionnaires 1) logged into the computer, 2) digitally scanned, 3) electronically checked for missing, contradictory, and impossible entries. 4) discrepancies sent to the site for resolution [1]
- Older trials on NIDA CTN used same process

How did you feel at 2:00 pm on Friday three weeks ago?

Bottom line – you can't clean paper questionnaire data after the fact.
You have one shot – point in time while the Patient is completing the questionnaire

Data Processing (cont.)

Paper questionnaires



Electronic questionnaires

Fraud

- Missed visit interview taker ....
- Diaries all in a row
- Cut & paste focus groups

Laziness

Causes – Aggravators:

- Unintentional incentives
- Lack of oversight
- Lack of system support (electronic data surveillance)

For Future Research

- Haven't seen a literature synthesis / review of accuracy on patient reported data
- Haven't seen a model / framework that practitioners can use to prospectively design data collection and processing for patient reported data

Take Aways

- Errors
  - exist, in fact are a force of nature
  - are dependent on source & processing
- As few steps as possible please !
- Count soup to nuts
- Know how clean your data are
- How clean is good enough – ask study statistician
- Dimensions other than accuracy come into play when others use your data

Questions & Discussion