



ECRI Institute's 24th Annual Conference on the Use of Evidence in Policy & Practice
Workflows, Workarounds, and Overworked Health Systems
George Washington University, Washington, DC
Tuesday, 28 November 2017 – 9:50a – 11:a

Rating the Raters: Measuring and Reporting Quality of Care

Brent C. James, M.D., M.Stat.
Quality Science

Disclosures

Neither I, Brent C. James, nor any family members, have any relevant financial relationships to be discussed, directly or indirectly, referred to or illustrated with or without recognition within the presentation.

IHI Global Trigger Tool

- Major teaching hospital; random sample w/ 325 adult inpatients; October 2004
- Record review performed March 21-22, 2005, by a team of 7 trained abstractors
- All charts, at all levels, reviewed twice

35.1% of all admissions had at least 1 care-associated event
26.0% had at least 1 event within index admission

(9.1% of all hospital admissions resulted from outpatient care-associated adverse events)

<u>Rate</u>	<u>Severity Level</u>	<u>Rate</u>	<u>Source</u>
53%	E - temporary harm, required intervention	52%	medications
33%	F - temporary harm, initial or prolonged hospitalization	20%	procedure complications
3%	G - permanent harm	13%	infections
7%	H - intervention required to sustain life	8%	care issues
1%	I - patient death	3%	device failures

Unusual findings: minimal issues relating to anticoagulants, insulin, and PCA pumps, which are much bigger at other institutions (LDSH has protocols in place for these). That yields an injury rate of 82 / 1000 patient days, while most other hospitals are just above 100 injuries per thousand patient days.

Extrapolating to a full year, about **132 'sentinel event' deaths** occurred.

Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, Whittington JC, Frankel A, Seger A, James BC. 'Global Trigger Tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Hlth Aff* 2011; 30(4):581-9 (April).

Implications for public reporting

IHI Global Trigger Tool found: **132** events

That year, that hospital reported: **9** events

All 73 hospitals in the region reported: **36** events

For-profit hospitals (*the “competition”*) reported: **0** events

Should the hospital deploy the IHI GTT?

Better event detection:

- Almost all care providers use *voluntary reporting* to find events
- On the evidence, voluntary reporting finds, at best, about 1 in 10 actual events (about 1 in 100 in typical use) compared to more sensitive methods (like the IHI GTT)

**~150 different groups rank hospitals.
They report *VERY* different results.
For example,**

*U.S. News & World Report “20 Honor Role” Hospitals:**

5 CMS Stars (2.4%):	1 hospital
4 CMS Stars (22.5%):	9 hospitals
3 CMS Stars (51.6%):	6 hospitals
2 CMS Stars (19.6%):	3 hospitals
1 CMS Star (3.9%):	0 hospitals

*U.S. News & World Report 27th Annual Report. One “20 Honor Role” hospital was not included in the CMS Star ratings.

Of 4,655 acute care hospitals, 1,428 (30.7%) were rated a “Top 100” hospital (top 2%) by at least one external ranking group.

*Paul Keckley while at Deloitte, ~2013

Core problem: measuring quality

1. ***Signal-to-noise ratios***
2. ***Aim defines the*** (measurement) ***system***

1. Signal-to-noise ratios

- *Walter Shewhart: All complex systems contain an element of random variation;*
- *measurement, itself, is a complex system.*

That leads to a key question:

***How much of measured variation
arises from the process itself,
as opposed to the measurement system
through which one views the process?***

A technical detail ...

Finding a statistically significant outlier means, exactly,

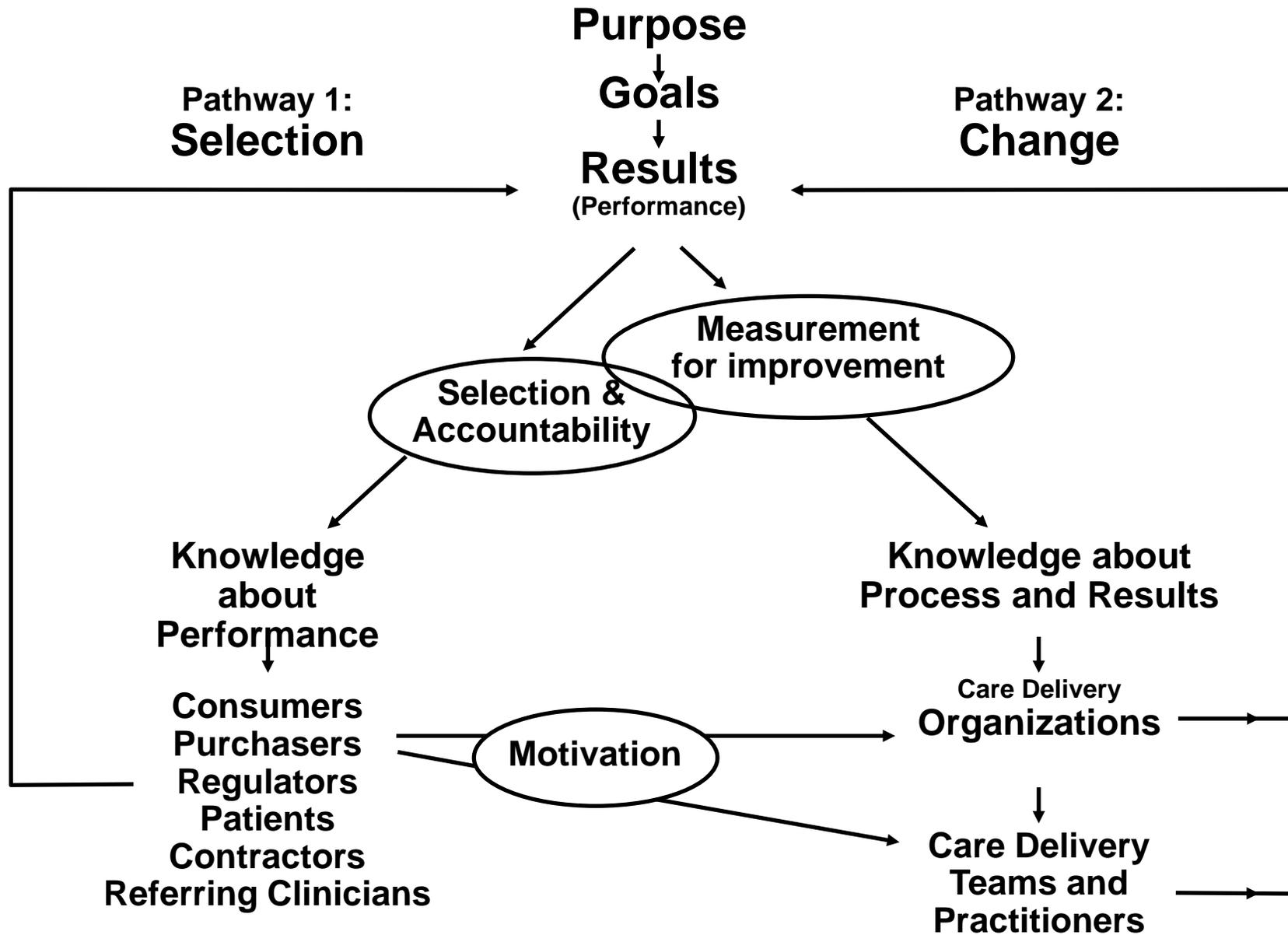
that it is possible to track the outlier back to a legitimate cause

But by itself,

the initial analysis does not identify the cause!

At one large integrated system, we:

- **Carefully designed measurement systems for clinical processes** (*58+ clinical registries; extending well beyond “available” data; and not relying primarily on secondary data, such as insurance claims*)
 - **Deployed them into clinical work flows;** (*Rule: avoid abstraction for routine data operations – it is expensive, introduces long delays, and often produces very high error rates*)
 - **Found significant numbers of statistical outliers; then**
 - **Tracked those outliers back to root causes**
- Between 30 and 50 percent initially tracked back to the data system, not the clinical process**



Ref: Berwick, D.M., James, B.C., and Coye, M. The connections between quality measurement and improvement. *Medical Care* 2003; 41(1):130-39 (Jan).

Selection approaches assume:

- 1. An ability to accurately rank** (*accurate ranking requires complete, accurate, unbiased measures*)
 - sufficient science (most of the critical risk factors are currently known)
 - find then accurately, completely assess all elements for all cases
 - reliable data extraction across disperse settings
 - high statistical resolution (mathematical problems w ranking)
 - methods to appropriately assign attribution
 - defensible methods to combine individual into summary scores
- 2. That consumers will respond to the rankings**
- 3. There is sufficient "good" system capacity**
within geographic reach, to handle the hoped-for concentrated volume
- 4. Poor performers will respond with real improvement**, *not just "better documentation," risk selection, or resource concentration*

Top-down selection measures

- ***Usually do not include all essential data elements*** (*entry, exclusion, & stratification criteria; clinical cofactors*);
- ***consume large amounts of resources***
through “after the fact” data abstraction;
- ***leaving no resources for actual performance management and improvement.***

Thus,

Selection measures, imposed in the name of accountability for quality, can block improvement and actively damage care quality

Measurement for Change / Learning

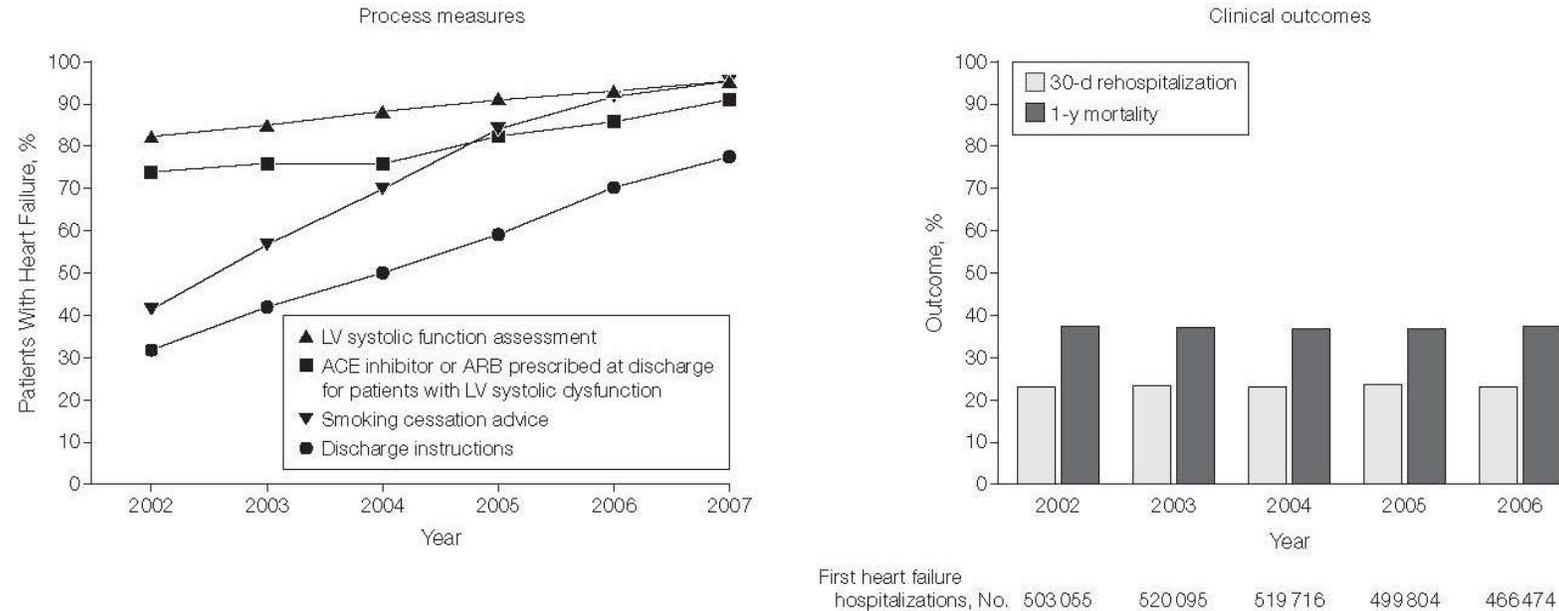
1. **Generates very different data sets**
 - strong, evidence-based method derived from RCT data system design
 - intermediate and final clinical, cost, and satisfaction outcomes
 - optimized for process management and improvement
 - more extensive, clinically focused than typical Selection measures
2. **Tolerates** (stable) **bias** – can still identify whether care is improving or getting worse
3. **Is parsimonious** (no “recreational data collection” while avoiding **availability bias**)
4. **Minimizes burden** - integrates into clinical workflow; tends to be what clinical teams must generate and use to deliver care
5. **"Contains" selection measures** - produces robust patient outcomes measures suitable for public accountability

Deming noted that

when people are pressured to meet an external target, they can:

- 1. Improve the system;**
- 2. Suboptimize the system** by “working harder:”
concentrate resources under the measurement spotlight, at the expense of work that is not under the spotlight = overall quality declines; or
- 3. Game the data,** a.k.a. “improve documentation”
 - *as pressure to “make the number” increases, reliance on methods #2 and #3 increases disproportionately; and*
 - *external pressure destroys internal motivation*

Figure. Temporal Trends in Process Measures and Clinical Outcomes for Medicare Patients Hospitalized With Heart Failure, 2002-2007



Source population for process measures comprised patients hospitalized at Joint Commission–accredited hospitals; source population for rehospitalization and mortality outcomes comprised Medicare beneficiaries 65 years or older; outcomes data for 2007 are not yet available. Beginning in 2005, the data collection method changed from the abstraction of randomly selected medical records for Medicare beneficiaries to the receipt of hospital self-reported data for all payer types. LV indicates left ventricular. Data from Curtis et al² and The Joint Commission 2008 quality report.⁶

Fonarow GC, Peterson ED. Heart failure performance measures and outcomes: Real or illusory gains. *JAMA* 2009; 302(7):792-4 (Aug 19).

Recent examples:

- ***VA waiting list scandal***
- ***No Child Left Behind testing scandal***
(with criminal prosecutions of school teachers in Atlanta)
- ***Wells Fargo Bank credit card scandal***

Given the complexity of clinical care,

any legitimate clinical outcomes measurement strategy will always include a mechanism for
(1) data system validation (audit)
and (2) data system feedback

Better has no limit ...

an old Yiddish proverb