

Predictive Modeling NEWS

Outcomes Measurement for Dummies ... and Smarties

by Al Lewis, President, Disease Management Purchasing Consortium Inc., Waltham, MA

Introduction

Healthcare and human resources executives who can distinguish between valid and invalid vendor proposals, outcomes reports and peer-reviewed articles will be able to allocate their resources effectively to undertake

population-based interventions. Virtually unknown 15 years ago, care management programs -- such as disease management, wellness and patient-centered medical homes -- have exploded in popularity in recent years. The idea driving all of these programs is that targeted use of additional workplace, ambulatory, physician extension and/or telephonic care resources can reduce hospital and emergency room usage among those with, or at high risk for, chronic disease and/or avoidable procedures, while improving the productivity of those resources.

Behind this idea is the assumption that, especially in the five most-lifestyle-related (or lifestyle-exacerbated) common chronic conditions -- asthma, coronary artery disease (CAD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD) and diabetes -- rates of hospitalization/emergency room (ER) use are high and growing, and therefore create a significant financial burden on the medical system. The Centers for Disease Control and Prevention (CDC) estimates that 75% of all spending goes for chronic disease, and that roughly 1 in 2 adults live with at least one chronic disease.

Most healthcare and human resources executives see a problem here, but the fact that healthcare costs keep rising unabated is circumstantial evidence of their collective inability to distinguish effective solutions to that problem from those solutions whose "effectiveness" is basely largely or totally on mistakes in the analysis, mistakes that go unnoticed by most readers, who assume that peer-reviewed reports or reports naming national health plans must have been checked and/or be generally correct in order to make it into print. That assumption ignores the key difference between care management outcomes studies and randomized double-blinded prospective control trials for pharmaceutical studies, the studies that most healthcare executives "grew up" on: The latter are conducted according to well-established regulations and all published statements need FDA approval. Having been conditioned to believe that what they read is generally accurate, healthcare executives and benefits managers tend to believe care management outcomes studies as well, even though they are totally unregulated. As a result, those outcomes are frequently implausible and often mathematically impossible.

Care management outcomes innumeracy isn't just of academic interest. Large amounts of resources are wasted on programs that don't work, driving up healthcare premiums. Public policies are made based on data that are obviously wrong, but that no one questions, driving up taxes. Paradoxically, the reverse problem is also true: Many decision-makers refuse to believe outcomes that are largely correct, mostly because they assume that all positive outcomes and promises are wrong. Hence, in many cases, initiatives that could really reduce costs are under-resourced. Two seemingly opposite resource allocation problems, one cause: innumeracy.

Innumeracy is not a "he said -- she said" issue. Math is immutable. A proof settles an issue. That's why this article can name names with no fear of lawsuits -- every statement in here is based on math. Nor does having actuaries attest to outcomes ensure numeracy. The mathematically impossible example in this article was perpetrated by Mercer Management Consulting⁵ the largest actuarial consulting firm, while Cigna, Kaiser and Wellpoint -- also associated with impossible or extremely implausible outcomes or promises -- employ large staffs of actuaries. Given the lack of rules, the inability of actuaries to understand care management outcomes and the general innumeracy displayed by many study sponsors, executives at all levels must rely only on themselves to determine whether the outcomes they are shown or promised are plausible.

The "Plausibility Rules"

Applying one or more of seven informal, common-sense rules can help determine whether a seemingly attractive outcome showing a high return on investment (ROI) is plausible. The rules are not footnoted or otherwise sourced, because there is no precedent and no governing body for validating care management outcomes. Instead, you must make your own decisions as to which rules to believe. The examples that will be shown, though, will support all of them.

Few readers will disbelieve even one of these common-sense rules, because plausibility analysis is sensitive, not specific. Its goal is not to validate every study that is truly valid, which would be a time- and expertise-intensive task, but rather to invalidate those claims that are almost certainly invalid, a first level of intellectual triage to avoid wasting time and, more importantly, to avoid making misguided resource allocation decisions. Even that seemingly modest level of intellectual triage will dramatically “thin the herd” of published outcomes, allowing executives to focus on evaluating interventions that might really work.

The Plausibility Rules are as follows, with their shorthand in parentheses:

- [1] Outcomes explicitly or implicitly cannot require cost declines exceeding 100% (the “100% Rule”).
- [2] Every element of resource use or group of people cannot decline in cost in programs aimed at prevention. In particular, the actual costs associated with prevention, such as primary care visits, drug use and testing, must rise (the “Differential Impact Rule”).
- [3] Declines in excess of 50% in any given resource category are much more likely to be the result of invalidity than of effectiveness (the “Reasonable Savings Rule”).
- [4] There must be a nexus between the goal of the program and the source of savings (the “Nexus Rule”).
- [5] Just as in pharmacology, where there is an observable time-dependent relationship between dose and response, cost cannot decline significantly faster or sooner than the related quality variables improve (the “Quality Dose-Cost Response Rule”).
- [6] “Control groups,” if not prospectively sorted into two like groups, based on objective data, before members are even contacted to determine willingness to enroll, are likely to mislead. This is especially true of “historic controls” (meaning pre-post), “matched controls” and “using the non-disease group as a control for the disease group” (the “Control Group Equivalency Rule”).
- [7] When one of those rules is violated, others are likely to be violated as well (the “Multiple Violations Rule”).

Examples of Plausibility: The 100% Rule

The 100% Rule is the only rule that is pure arithmetic. The limitation of percentage declines to 100% is a very elementary concept. And yet the level of discourse in the care management field is such that there are many explicit examples of its violation¹, of which two – that I know of -- have been packaged into marketing materials and onto websites and distributed. The first example: The last line of the following excerpt from the case study used in marketing claims a “230% reduction” in extended illness benefits among wellness program participants.

Wellness Program Case Study
The Children’s Hospital

The Children’s Hospital of Denver (TCH) started their first comprehensive wellness program in 2007, implementing a personalized approach focused around a high trust, high engagement strategy with US Corporate Wellness. The following provides data resulting directly from this program.

Access and Participation
All benefit eligible employees at TCH - approximately 3,200 people - were granted access to participate in the program. Those receiving benefits – approximately 2,400 people – were provided a moderate incentive to participate. The program grew quickly to 1,400 active participants upon launch and has since exceeded 2,000 active participants at the end of 2009. This comprised 63% of total eligible employees and a full **83% of incentivized employees**. Active participation was not based simply on completion of an online document or logging onto a site to register. Rather, it involved ongoing discussions with the individual’s own personal wellness coach in setting goals, identifying areas for change, etc.

Sick Time
Like many organizations, TCH combines the first 16 hours of sick time along with vacation into a PTO bank. Hours tied to sick time beyond 16 hours are categorized as EIB (Extended Illness Benefit), and this was the focus of the analysis. Maternity and anomalies (totaling 3%) were pulled out and the remaining 97% plus of the total population was analyzed at TCH, with the following results:

- Wellness program participants are **230% less likely to utilize EIB** than non-participants

The second example is also from the realm of absenteeism reduction. On its website, TrestleTree, which calls itself a “health transformation organization working to enhance healthy behaviors of employees and their families,” at one point boasted that “TrestleTree’s participants showed a 300% reduction in illness-related work absenteeism last year.” It has subsequently been removed from the vendor website following a (surprisingly lengthy) impromptu remedial tutorial on how percentages work.

Below is an unedited text excerpt from a slide presentation² in which savings of \$5,000/person/year (net savings, meaning after fees are subtracted) generated by a care management program for commercially insured members was claimed, where this number was said to be for the “average” person. However, the average person doesn’t even incur \$5,000/year in paid claims – and certainly not in claims that could be considered even theoretically avoidable -- making it impossible to reduce claims by this amount, especially net of fees, a clear violation of the 100% Rule that the presenter acknowledged after this was pointed out.

An example like this demonstrates the need for more instruction in the health outcomes numeracy field, both in general and also specifically because the presenter runs a non-profit called The Center for Health Value Innovation (CHVI), whose “Vision Statement” says that “[CHVI] will be the trusted resource to demonstrate how engagement in health can improve accountability and economic performance,” and yet was unable to recognize on her own that it is not possible to save \$5,000/year/person in a commercial population, as the slide excerpt below shows.

<p><u>Examples From the Literature</u></p> <p>Care Delivery</p> <ul style="list-style-type: none"> ▪ Significant improvement in absenteeism rates ▪ Net savings per program participant of over \$5,000 ▪ Significant improvements in guideline adherence for preventative and care management services such as mammography ▪ Overall downward shift in risk, including lower cholesterol, improvements in diet and better management of blood pressure
--

Invalidity in the Public Sector

To support the thesis that policymakers are equally innumerate – and therefore that the country’s tax dollars are being misallocated – we must turn to the public sector, where examples of innumeracy abound.

North Carolina Medicaid Patient-Centered Medical Home

Let us consider the much-publicized patient-centered medical home in the North Carolina Medicaid program, the poster state for the Multiple Violations Rule, having violated at least five of the six other rules. (It probably also violated the Quality Dose-Cost Response Rule, but the HEDIS indicators for the state for the baseline period were not readily available for comparison to the more recent HEDIS indicators, which were below average in the relevant categories.)

The main finding was that the state saved \$284 million to \$314 million in fiscal 2006, the last year studied, versus prediction. The outcomes – the cost savings allegedly realized by the most recent available year studied (2006) as compared to the trended baseline period (2000 to 2002) -- were analyzed two different ways by Mercer Consulting. The first was by the resource category:

AFDC - All Rate Cells Combined			
SFY06 Member Months	7,962,681		
Category of Service	SFY06 Projected Benchmark PMPM ¹	Actual SFY06 PMPM ²	Estimated Savings from Benchmark ³
Inpatient	\$ 43.25	\$ 23.16	\$ 159,963,111
Outpatient	\$ 23.47	\$ 17.73	\$ 45,660,400
Emergency Room	\$ 15.11	\$ 11.30	\$ 30,324,253
Primary Care, Specialist	\$ 56.90	\$ 50.91	\$ 47,751,911
Pharmacy	\$ 31.72	\$ 30.14	\$ 12,601,550
Other	\$ 30.78	\$ 30.46	\$ 2,516,055
Totals	\$ 201.23	\$ 163.70	\$ 298,817,281

Here are the questions to ask, based on that analysis:

- [1] Why are the doctors so supportive if they're working harder and making less money?
- [2] Why did Mercer lump specialists and PCPs together, even though the latter were supposed to replace the former?
- [3] How can the doctors be putting more people on preventive meds if the drug expense went down? (Note that generic substitution is not a part of medical homes. This is supposed to be savings due to medical homes.)
- [4] Where are the patients getting their care, if not from hospitals, ER visits, OP clinics, doctors, drugs or "other"?

The answer: Never, ever believe a result in which all resource uses decline. People have to be getting their care from somewhere, so the predicted what-would-have-been is incorrect. When you are looking to determine if the predicted number was wrong, you focus on the biggest number. If that number is wrong, so are all the others.³

Note that every category of utilization fell, a Differential Impact Rule violation. The assumption behind that rule is that people have to be getting their care from somewhere. Absent a major commitment to an expensive and effective wellness program, the need for care will not disappear. If people aren't getting care from the hospital, the emergency room or any other outpatient setting, they are getting their care from doctors and adhering more to their medications – meaning, obtaining more preventive drugs. But in North Carolina, the cost of doctors and medications is shown as declining too. (Unit cost cannot be responsible for the decline, since this type of care program does not involve generic substitution or change contracted rates, except to raise them for primary care providers.)

One curious feature of the alleged physician cost decline is that even though one goal of patient-centered medical homes – in the analysis, "medical home involved every non-disabled Medicaid member" -- is to substitute timely and effective primary care for specialist care, the two categories are reported together.

A cynic might infer that the reason for combining those two categories into one must have been that specialist visits did not decline relative to primary care visits. If that cynic did a Freedom of Information Act request asking for primary and specialist visit numbers as this cynic did, that cynic would find support for that theory.

The state also provided the outcomes by age and gender, as follows:

State of North Carolina

**Attachment 6
SFY06 Savings
Using Statewide Benchmark
(by Rate Cell)**

Final

AFDC - Specified Categories of Service				
Age and Sex Description	Member Months ¹	SFY06 Projected Benchmark PMPM ²	Actual SFY06 PMPM ³	Estimated Savings from Benchmark ⁴
< 1 year M & F	670,070	\$ 411.38	\$ 186.80	\$ 150,479,255
1 - 13 years M & F	4,672,745	\$ 102.70	\$ 100.37	\$ 10,901,303
14 - 18 years F	596,909	\$ 224.57	\$ 166.58	\$ 34,614,787
14 - 18 years M	547,434	\$ 112.82	\$ 109.84	\$ 1,632,831
19 - 44 years F	1,167,464	\$ 413.69	\$ 359.99	\$ 62,695,031
19 - 44 years M	174,219	\$ 452.90	\$ 310.30	\$ 24,844,077
45 years & up M & F	133,840	\$ 665.60	\$ 563.62	\$ 13,649,997
Totals	7,962,681	\$ 201.23	\$ 163.70	\$ 298,817,281

1 - CCNC/ACCESS only member months for SFY06.

2 - The Statewide Benchmark SFY06 PMPM was calculated using the historical 36 months of data from SFY00, SFY01, and SFY02. The PMPM shown here is calculated by weighting each rate cell's SFY06 base PMPM with the actual CCNC/ACCESS member months distribution by rate cell for SFY06.

3 - Calculated using the date of service data for SFY06; represents all CCNC/ACCESS program (I, II, and III) costs for dates of service from July 2005 through June 2006.

4 - Projected savings calculated using the SFY06 actuals; the benchmark minus the actual, multiplied by the actual SFY06 CCNC/ACCESS member months, equals the projected savings.

The 54% savings in the “<1 year M & F” line violates both the Reasonable Savings and Nexus Rules, because of the implausible magnitude of the savings and the overwhelming unlikelihood that a program geared towards controlling chronic disease would show massive savings, or any noticeable savings, in babies, the age category least likely to have chronic disease. Additionally, the magnitude of the savings, plus the Multiple Violations Rule, points to the 100% Rule as potentially being violated.

The analysis proceeds as follows.

There are four major components of cost in this age group.

- [1] physician costs;
- [2] normal delivery costs;
- [3] drug/other; and
- [4] neonatal care.

While the data weren’t provided, pediatrician costs surely increased because primary care doctors receive a higher capitation rate to participate in a PCMH model. Normal delivery length of stay (and hence presumably costs per normal delivery) climbed about 5% over the period.

Though not detailed separately, drug and other costs, part of preventive care, can be assumed to also rise, or at least not fall, in a model oriented towards prevention. That leaves inpatient neonatal care costs, which may approximate half of all healthcare costs for the 0-to-1-year-old group, as the only possible source of savings in this age category.

To achieve a 54% overall reduction in this age category, neonatal utilization would have to drop by more than 100%, since if neonatal utilization represented half of all costs in the category and fell by only 100%, total costs would fall by only 50%, rather than the claimed 54%.

(The actual percentage of total cost in the <1-year-old category accounted for by neonates was not disclosed, and in the United States as a whole is not tracked. “Half” is used as a common-sense substitute for a true number, but it does not matter what the true number is.

The actual change in neonatal resource usage may be independently verified using a free, federally supported database that captures nearly every hospital discharge in the country. The discharges and days for neonates are compared to the discharges and days for normal deliveries. The three years of baseline (2000 to 2002) are contrasted to the study year, 2006. The line in bold italics shows that neonatal utilization fell by 1%, not 100%, over that period – a mistake of two orders of magnitude.⁴ (The state and its consultants would no doubt argue that neonates would have risen dramatically over that period without the PCMH, but a similar analysis for South Carolina would show that – even absent a PCMH – neonatal utilization rate fell by almost the same amount over the same period.)

What is a typical distribution of medical spending by resource component?

At several points in this article, I make reference to certain resources, like hospital and pharmacy costs, as a percentage of total medical spending at a health plan. As a general rule of thumb – and plausibility is all about rules of thumb, because precision and accuracy are often inversely related – figure that hospital costs are 40%, physician and drug costs are both 20% and everything else sums to 20%. Of course, this will vary by condition, federal/state/private payer type and severity. For instance, greater illness burden is associated with a higher proportion of inpatient costs.

DRG		2000	2001	2002	2006
386-390	Non-normal discharges	33,631	30,227	27,776	32,390
	LOS (length of stay), days (mean)	6.4	6.9	7.1	7.1
	Discharge days	216,257	207,897	196,181	229,969
Diagnosis Related Group 391, Normal Newborn					
391	Total number of discharges	79,875	80,419	81,090	89,643
	LOS (length of stay), days (mean)	2.0	2.0	2.0	2.1
		159,750	160,838	162,180	188,250
	Total newborns	113,506	110,646	108,866	122,033
	% Non-normal discharges	29.6%	27.3%	25.5%	26.5%
	% Normal discharges	70.4%	72.7%	74.5%	73.5%

State Medicaid programs generally have a stake in sustaining high savings projections in programs like North Carolina's because the money they spend on those care management programs is largely financed by the federal government through "matching," so they often seek and obtain "validation" of their programs rather than objective review. A simple remedy to this moral hazard would have been to require that, rather than Medicaid itself, an independent state agency review savings findings, but this remedy was not part of the comprehensive health reform package.

Contact Lewis at 781-856-3962 or at alewis@dismgmt.com.

FOOTNOTES

¹ When President Obama was campaigning for healthcare reform at an Ohio rally, he received his loudest applause for stating that healthcare reform would "reduce health insurance premiums by 3,000%."

² The presentation was given at the Care Continuum Alliance's 2010 Annual Leadership Forum. However, the Care Continuum, like virtually all organizations that put on conferences, does not take responsibility for the speakers' content once they have been accepted to speak (even though a co-presenter was an employee of the Care Continuum Alliance). To do so would jeopardize an organization's ability to offer continuing education credits.

³ One comment on www.thehealthcareblog.com suggested that perhaps every element of cost declined because more preventive care was being rendered. However, while insulating your house may save money, it does not reduce your cost of insulation, and the cost of preventive care appeared to have declined. Additionally, it turned out that preventive care probably did not improve, as North Carolina's obesity and diabetes rate both rose more than the national average during the last decade, as reported by the Kaiser Family Foundation.

⁴ In all fairness to the state, there was a slight mix change from extremely premature babies to less-premature babies, but this mix change, even when paid for by a severity-weighted DRG as opposed to a per diem, would only change the overall result by a few percentage points. Likewise, in all fairness to the state, the federal database is all-payer and a sizable minority of neonates are not Medicaid. However, there is no *a priori* reason to think that non-Medicaid neonates would have jumped to make up for a theoretical massive decline in Medicaid neonates.

⁵ Population Health Management, Volume 11, Number 5, 2008, Testing the DMAA's Recommendations for Disease