

Effective Application of Comparative Utilization Data in Managed Care Organizations

R. Curt Hatch, CFA, MBA

Summary

The most critical determinant of managed care organizations (MCOs) success is their ability to capture and utilize *data*. Understanding—and shaping—the factors that affect MCOs' bottom lines requires a multi-dimensional approach to data management. A vital component of this approach, and one that too few organizations possess, is the effective application of *comparative utilization data*. The ability to acquire external utilization comparisons, and use them to identify cost-saving opportunities, is an essential survival skill in the management of an at-risk medical budget.

The American Association of Integrated Healthcare Delivery Systems (AAIHDS) is considering establishing an actuarial data repository as a service to its members. This project would involve the collection of “claims-level” data from at-risk MCOs nationwide. The data would be mapped into standardized schema, ensuring consistent processing of all types of claims, enrollment, capitation, and related data. The benefits of this centralized warehouse would then be available to association members through a) regular side-by-side comparisons of participants' utilization and costs against the balance of the data universe, and b) customized studies of utilization management (UM), outcomes, disease management, and other population-based managed care studies.

Key Points

- ❑ Comparative utilization data offers MCOs critical insights into cost-containment measures and increased efficiencies.
- ❑ By overcoming the barriers to external benchmarking (e.g. conformity, comparability, and cost) MCOs can effectively integrate comparative data into their UM processes.

FOR YEARS, MCOs HAVE STRUGGLED to make effective use of the volumes of data they have at their disposal. Many have discovered the importance of this capability only when it was too late. The rash of IPA failures in California, Texas and other markets was caused in no small part by their inability to identify adverse utilization trends before they became irreversible. Managing insurance risk is a highly leveraged business, where relatively modest fluctuations in cost can have catastrophic effects on the bottom line.

The core of good data management is standardized reporting. Every MCO should have a consistent library of periodic reports that will highlight trends as they begin to emerge and enable the MCO to track trends already on the “radar screen.” The MCO should include a variety

of report categories that capture financial and utilization statistics—from both a static and longitudinal perspective—of its population. The database also should have discursive or drill-down capabilities, so that an issue from one report can be examined in further detail on another. It is critical to note here that any use of comparative data is impossible if the MCO cannot capture information on its own population and providers.

Effective data use also means presenting it in a way that provides the viewer with some level of insight. Abstract figures (“Our commercial hysterectomy rate is 3.6/1,000.”) are seldom useful unless they are compared to some other metric. *Comparisons create context*. If the MCO knew that last year's hysterectomy rate per thousand was 2.5, or that the national average is

2.1, the figure begins to assume some significance. Whether internal (among physicians, across time, between contracts, etc.) or external (to other organizations); successful MCOs employ this model for “comparative context” throughout their data management structure. An important component of this comparative data model is external benchmarking.

Comparative/Benchmarking Data

The old saw about healthcare being a local phenomenon (“if you’ve seen one market...”) is undoubtedly the most overused in the industry. But it does have an important corollary for managed care. Local markets develop peculiar practice patterns that, more often than not, deviate from the courses of both quality and efficiency. How can an MCO address these patterns if it is navigating *only* by its own rear view mirror?

Clearly, it cannot. MCO managers must be able to make valid comparisons between their internal utilization patterns and accepted benchmarks from other communities. Otherwise, undesirable patterns reinforce themselves and become increasingly “normal.” The obstacles to cross-market comparisons are numerous, and include

- conformity in data-collection techniques
- comparability between the subject and the benchmark data
- cost

The first barrier—*conformity*—is typically the most daunting. There are initiatives throughout the healthcare industry aimed at compiling useful benchmark data. Most of them are survey-based, where a respondent is asked a question such as “What was your \$PMPM cost for orthopedics last year?” On its face, this is a fairly specific question. But the ambiguity inherent in this—and any other managed care data question—invariably compromises both the credibility and the usability of these benchmarking surveys. Consider some of the qualifying questions that one might ask when trying to respond to a typical utilization survey (assuming, rather naively, that the respondents are highly motivated to produce the most accurate figures they can):

- The survey asked for the total number of knee arthroplasties? How do I count the assistant surgeon claims that came in under the same CPT code? When one patient has both knees*

done, is that a frequency of one or two? Should point-of-service be included in commercial data?

- Cardiology costs \$PMPM? We sub-capitate invasive cardiology, should that be included? What about the capitation; do you want the FFS-equivalent cost of the services, or just the cap \$ paid out? What about the EKGs done by our internists who have cardiology subspecialties? Would you include heart transplants that aren’t included on our side of the division of responsibilities with the plan? We also get split-billed for some echos on a FFS-basis, is it OK that we just ignore the modifiers?*
- No problem, we report bed-days per thousand every month. But does that include psych? SNF? Chemical dependency? Observation days? And our calculation places the entire length of stay in the reporting period that the admit occurs—is that okay?*

These hypothetical questions illustrate the vagaries of collecting data through a survey process. Professionals who have tried to apply survey data to their UM process has seen how these problems can discredit the data. Therefore, most managed care practitioners look skeptically on survey data for just these reasons. The bottom line is that no data is going to be terribly useful unless it contains claims-level detail, and has been carefully scrubbed and mapped to ensure consistency.

The second obstacle is *comparability*. “Data from other markets does not apply here in My City, U.S.A.” Physicians’ resistance to considering data from other locales stems from several core concerns—some justified, and some less so. The notion that utilization patterns in a local area might be unique is not, in itself, problematic. What matters is whether the uniqueness is in some way explainable, and whether the explanation is consistent with efficient, high-quality care. The rationale that “we have higher imaging costs in Schmeckler County because we have seven outpatient CT scanners serving our 14,000 patients” should not suffice.

The source of any comparative actuarial data should be understood to as great an extent as possible. The more that is known about its composition, the greater its potential to shed light on the MCO’s data by comparison. If the subject

IPA operates in an advanced managed care market, there is considerable risk in comparing it with an unmanaged indemnity population. Even within managed care markets, there are highly advanced (i.e., utilization has been substantially reduced) and relatively primitive markets. MCOs that make extensive use of capitation will tend to have lower utilization experience (missing encounter data notwithstanding) than those that are fee-for-service. These dissimilar payment models don't necessarily render the comparison invalid, but do suggest a potential reason for any differences. Remember as well, that medicine is a rapidly changing field. Two-year-old use rates may be too old in the case of colonoscopies.

Finally, comparing cost (\$PMPM) is always dangerous. Cost involves the confluence of two primary variables—use rates, and cost/unit (*see Exhibit 1 at end*). It's wise to be leery of comparisons that cannot “decompose” these two variables and peg the difference to one, the other, or both.

In fact, managing \$/unit—a contracting function—should not be confused with utilization management. When drawing comparisons between markets, the focus should be on use rates. The volume of services that a population requires should be fairly predictable from region to region. Pricing, however, is highly variable, and driven by local supply, demand, geographic layout, competition, negotiating clout, consolidation, and the like.

Finally, there is *cost*. Indeed there are some actuaries out there who can produce data that deals adequately with the first two issues, and produce valid, internally consistent comparisons. But the cost of employing their data throughout the scope of the UM process can be prohibitive. Costs could easily run \$50,000 to 100,000+ on an annual basis. Very few organizations can afford to acquire such costly intellectual property, regardless of its value.

Using Comparable Data: Lessons from the Field

Physicians are extremely wary of statistics that try to reduce the art that they practice to a purely quantitative equation. And rightfully so. There are very real limitations on the ability of statistical measures to capture the subtleties of clinical care processes. The key to obtaining buy-

in is to recognize these limitations early and often. Above all else, employing benchmarking data is a *process*. It cannot happen overnight. Any organization that intends to accomplish any meaningful changes in behavior inside of a year is bound for disappointment. The focus should be on building a collaborative effort with physicians, in an environment of openness and trust. It should also be conducted with certain parameters in mind:

- 1. Start with good data.** Many organizations begin the benchmarking battle by falling directly on their own swords. The quickest way to do this is to begin with bad data. If the benchmarking data does not meet the conformity and comparability tests referenced earlier—do not try to use it. It should also be exposed to whatever reasonableness and completeness tests that can be concocted. Physicians are extremely adept at finding weaknesses in data. If the MCO's data is not iron-clad, any vulnerability will be quickly discovered.

- 2. Treat benchmark data with care.** Another imperative is to begin the process with the right message. The organization should not declare the benchmark data as an end-all and be-all. It should be made clear that external data has one, and only one purpose. It is to be used to “uncover clues that might lead to utilization management opportunities.” It should never be viewed as prima fascia evidence that someone is wasting money. The investigation begins with the identification of a variance (between the benchmark data and the subject).

- 3. Critique conclusions mercilessly.** As a general rule, *all variations should be analyzed*. The analyst's goal should be to find a plausible reason for any abnormal cost/utilization pattern. Often, the underlying variations can be quickly attributed to data errors, fee schedule differences, random variation, population differences, regional health trends, benefit differences, DFR (division of financial responsibility) differences and so forth. Many explanations represent the terminus of the analytical trail; others point to continued analysis or a need for change.

- 4. Lay aside the “big stick.”** One of the most crucial errors (and one that is tempting to many evolving MCOs) is to attach financial consequences to benchmarked performance. Often, they will start by tying bonuses or other

contingent monies to a specialty \$PMPM target established through an actuary or independent survey source. This strategy can be a recipe for disaster. Seldom are the benchmarks sufficiently evaluated to ensure that they are appropriate. Every specialty has its own special situations, nuances, and related data issues. Without months of analysis and discussion, they will defy any blanket comparison to a standard. It deserves repeating once again that bad data will destroy any benchmarking initiative. Furthermore, physicians will be much more receptive to an open dialogue if it does not have the potential to skewer them financially.

5. Involve physicians at every level.

Most physicians operate under the premise that no other community is like theirs. They dismiss any attempt to contrast or compare. However, their involvement is essential. Without it, benchmarking—and UM for that matter—would be simple. The reason that most organizations either fail to make changes or give up trying to manage their underlying use rates is because this is where the “heavy lifting” takes place. The more meetings, the more communication that can take place, the greater the odds of successful intervention.

6. Screen initiatives carefully. A successful benchmarking effort can filter a broad range of potential targets, but should ultimately focus on a relative few. Any intervention (e.g., education, financial penalties, guideline development, pre-authorization) should be preceded by an evaluation of some discriminating questions:

- *Significance:* Is there sufficient financial opportunity to justify the effort?
- *Actionability:* How likely is it that the potential intervention will be successful?
- *Precedence:* Will these actions send an important message to providers?
- *Related Issues:* Are there other considerations (especially political) that advise caution?

At some point in every benchmarking investigation, there is a pivotal decision point. This is the point at which the remaining hypotheses coalesce around a finding, and a decision must be made on whether or not to proceed. When the analysis is complete, and differences still exist, that decision should again

hinge on the *reason* for these differences. To put it succinctly, some warrant action and some do not. Exhibit 2 shows proposed general guidelines to frame this decision.

Issues in the latter category (Opportunities for Management) beg for further attention. But physicians will often resist. Frequently, this reaction is a matter of turf protection. It can also be exacerbated by the fact that external data is brandished against them like a weapon. But the argument will still center on the reliability of the data, and of benchmarking efforts in general.

MCO managers must work diligently to break down these arguments. No two markets are identical. But most enjoy more similarities than differences. Specialists may employ one procedure in lieu of another in this market, however, the difference should still be noted, measured, and assessed based on its contribution to quality and efficiency. It stands to reason that there is generally one best path to treating a given condition. As expert knowledge bases (including evidence-based recommendations and guidelines) become increasingly available to practitioners, the variations in care—through adoption of best practices—will narrow. The value in making comparisons now is to force the dialogue about what approach is best.

Applied Examples of Benchmarking

A number of real-world examples illustrate the benchmarking process in action.

Benchmarking Case 1

Specialty: OB/Gyn

Initial Presentation

Hospital data showed that mean LOS was longer than benchmark for abdominal hysterectomy (3.5 vs. 2.5 for the comparable group).

Analysis

- Analysis of source data confirmed applicability of data.
- Individual physician distributions showed three outliers whose LOS routinely went three rather than two days.
- A specialty group meeting was organized to explore potential causes.

- ❑ The first group meeting immediately focused on surgical strategies, in which it was clear that several physicians were using a more expeditious suturing/closing technique to economize on surgery time. They were under the impression that the driving cost consideration was time in the OR. The medical director informed them that the hospital contract was a per diem arrangement, and that an alternative (subcuticular) closing technique could consistently reduce length of stay by one day.

Conclusion

- ❑ Two working sessions aimed at educating the participating providers on the economics of this decision were organized and conducted.

Benchmarking Case 2

Specialty: Gastroenterology

Initial Presentation

Colonoscopy utilization rates (commercial) appeared dramatically higher than the benchmark data (41.4/000 vs. 18.6/000 for the comparable group).

Analysis

- ❑ Initial exploration of the topic with the GI group attributed the contrast to the use of colonoscopy as primary screening for colon cancer.
- ❑ Detailed, joint review of case lists divulged the fact that certain “home-grown” CPT codes were being included in the colonoscopy volume, that were skewing the comparison. The subject use rate was revised downward to 39.4/000 (not material).
- ❑ Guidelines from ACG, ACC and other sources were compiled to verify national support for subject protocols. Endoscopic volume by age group was reviewed carefully to confirm guideline compliance. Concurrent reductions in flexible sigmoidoscopy and hemoccult studies were identified.

Conclusion

- ❑ Quality of care objectives were deemed paramount, leading the utilization

management committee to endorse current utilization patterns.

- ❑ Subsequent focus moved toward redirection of technical component, since two facilities among the group’s contracting alternatives had rates that were \$200/unit lower than competing facilities.

Benchmarking Case 3

Specialty: Dermatology

Initial Presentation

Commercial \$PMPM dermatology costs were nearly triple the benchmark data (\$1.52 vs. \$0.56 for the comparable group).

Analysis

- ❑ \$PMPM costs were disaggregated into fee schedule and use rate components.
- ❑ Variances were driven almost exclusively by use rates, with pronounced deviations in the areas of lesion excision, skin/tissue rearrangement, surgical pathology, and lesion destruction.
- ❑ Since virtually none of the pathology work related to dermatology was referred out to pathologists, an allowance was made for this “scope of service” difference. Further analysis concluded that a reduction of approximately \$0.18 PMPM was associated with this adjustment. This reduction still left considerable unexplained variation in costs.
- ❑ Cost/referral analysis disclosed a subgroup of providers in a single practice whose \$/referral and RVUs/referral were two to three times those of both the normative data and the balance of the derm panel.
- ❑ The data was shared with all members of the derm specialty panel.
- ❑ Representatives of the outlier group consistently refused invitations to participate in a dialogue around this issue.

Conclusion

Since the utilization issues were so significant and intractable, and the IPA had ample coverage from other providers, the initiative eventually led to a termination of the subgroup’s contract. Volume was then redirected to other dermatologists in the

community with no interruption in care or patient access.

Benchmarking Case 4

Specialty: Orthopedics/Radiology

Initial Presentation

Certain MRI procedures, namely those relating to lower leg, appeared to be utilized at higher rates than the benchmark data (roughly 70 percent higher across key MRI CPT codes). See Exhibit 3.

Analysis

- ❑ Initial discussion with orthopedic surgeons provided tentative validation that MRI was being rather aggressively employed in treating knee pain. Justifications included the rationale that the local community was highly sophisticated and consistently demanded the best technology available.
- ❑ Specialist-specific analysis did not disclose any significant outliers among the orthopedic panel.
- ❑ National guidelines for both the treatment of knee pain, and the use of MRI in the diagnosis of osteoarthritis and other knee disorders were collected and reviewed by both the medical director and a leading physician from the orthopedic group.
- ❑ Using a variety screening techniques, data correlating specific imaging procedures with related knee pain episodes were gathered and compared with several of the guidelines.
- ❑ Local variations were noted and were deemed material.

Conclusion

- ❑ MRI, as a modality, was reintroduced to the MCO's pre-authorization list. Specific clinical criteria were circulated to the authorization techs and to the affected physicians.
- ❑ Ongoing education efforts were deployed in the PCP, orthopedic, and radiology care settings.
- ❑ MRI utilization (perhaps as a result of the "big brother" effect) did stabilize and even decline in the ensuing two quarters.

Future of Benchmarking

External benchmarking data can be a powerful tool for identifying utilization management opportunities. But there are numerous barriers to effective exploitation. Many MCOs that have attempted to build these initiatives have met with minimal success, but the problems they encounter generally are a function of execution rather than a flaw in the concept itself.

The biggest challenge is gaining access to reliable, high-quality, benchmarking data. This problem must be solved before benchmarking can be widely employed by the nation's MCOs.

Nevertheless, the steps to successful implementation are not complex. These simple principles are variations on the themes that drive success in other aspects of the UM process as well. They include:

- ❑ Commitment/follow-through: Physicians and other key constituents must perceive that the MCO is committed to these initiatives.
- ❑ Resources: Personnel, information technology, and funding are essential to feeding this effort.
- ❑ Data and analysis: Combine rigorously scrutinized data with constant follow-up, reevaluation, and flexibility
- ❑ Process: Ensure that the people who will be affected by an initiative are meaningfully engaged at all phases of its evolution.

R. Curt Hatch, CFA, MBA, is president of ManagedCare.com, a reporting and data management firm with IPA, HMO, and IDN clients across the U.S. He is also executive director of the Mesa County Physicians IPA, a 210-physician medical group with 50,000 at-risk enrollees in the state of Colorado.

EXHIBITS

Exhibit 1: Separating Use Rates and \$/Unit

MANAGEDCARE.COM

**Local MCO
Commercial
Utilization By-Procedure Grouping Comparable (F-05C)
December, 2000 - November, 2001**

Parameters:

Claims entered through:	February 26, 2002	Claims Included:	Paid and approved	MOC reference:	MCP01-0200-1
Period Selection:	Date of service	Date prepared:	March 27, 2002		

Principal Specialty	CPT Group Description	Procedure Code Range	Subject Data		MOC Comparable Data			Rate/1,000 % Var to Mean	
			Rate/ 1,000	\$/Unit	Low	Mean	High		
Allergy									
	Allergy Testing	95004-95078	282.6	4.41	79.9	165.1	364.7	4.37	71%
	Allergen Immunotherapy	95115-95199	297.5	12.37	68.1	165.4	533.6	10.65	80%
Cardiology									
	Other Therapeutic Services (incl. Cardioversion, CPR,	92950-92979	0.8	267.63	0.2	0.6	1.7	189.89	46%
	Percutaneous Coronary Interventions	92980-92998	1.0	1,830.31	0.6	1.0	2.0	956.26	-3%
	Intracardiac Electrophysiological Procedures	93600-93662	1.7	1,278.60	0.2	1.1	1.7	649.82	53%
	Echography - Heart	I3C	40.5	115.46	21.6	34.4	55.2	114.88	18%
	Echography - Carotid Arteries	I3D	2.7	112.56	0.8	2.5	3.9	124.82	8%
	Cardiac Cath incl. Imaging/Procedure	I4A	15.7	190.66	5.8	11.3	23.6	211.56	36%
	Pacemaker Insertion/Removal	P2E	0.2	1,032.14	0.1	0.2	0.6	534.89	1%
	Electrocardiograms (routine and rhythm)	T2A	94.9	32.54	53.1	72.1	151.1	26.66	32%
	Cardiovascular Stress Tests (incl. Echo)	T2B	31.7	121.45	7.2	24.2	51.0	81.72	31%
	ECG Monitoring	T2C	9.2	101.28	0.7	6.0	16.1	100.70	53%

Exhibit 2: Evaluating Utilization Variances

Explanations That May Conclude the Analytical Process	
Explanation	Example
>Epidemiology	>Higher skin cancer rates for a plan operating in Hawaii
>Provider scope of service	>Lower OB costs where family practitioners are delivering babies, or higher orthopedic costs where orthopedic surgeons use in-house, rather than hospital imaging
>Statistical insignificance	>The number of endarterectomies in a group of 6,200 commercial members
>Variances involving non-discretionary or essential services	>Radiation therapy treatments for prostate cancer patients

Explanations That Suggest Opportunities for Management

Explanation	Example
>Service supply/availability	>Lab costs in an IPA, where many of the PCPs have in-office labs
>Aberrations tied to only a subset of providers	>The utilization pattern is inconsistent, driven by a minority of the providers in that area
>Motivated by financial incentives	>Utilization of lesion excision codes increases after they are removed from the PCP cap and paid FFS
>Inexplicable variations	>If no explanation can be found—conclude that the service may be over-utilized

Exhibit 3: MRI Use Rate Data

Proc Code	Description	Subject Group			MCC Comparable Data				
		Rank	Use Rate/1,000		Rank	Use Rate/1,000			
			PCP	SCP	Total		PCP	SCP	Total
99213	OFFICE/OUTPATIENT VISIT, EST	1	1,098.6	296.6	1,395.2	1	1,059.3	242.1	1,301.4
99214	OFFICE/OUTPATIENT VISIT, EST	2	181.9	184.7	366.6	2	267.8	134.2	401.9
59400	OBSTETRICAL CARE	3	0.0	11.8	11.8	3	0.4	10.4	10.8
99244	OFFICE CONSULTATION	4	4.4	78.0	82.4	4	2.3	48.7	50.9
88305	TISSUE EXAM BY PATHOLOGIST	5	0.1	138.3	138.4	7	0.0	84.8	84.8
99215	OFFICE/OUTPATIENT VISIT, EST	6	82.3	25.8	108.1	6	40.1	27.7	67.7
99243	OFFICE CONSULTATION	7	2.1	90.2	92.3	5	1.9	63.5	65.4
97110	THERAPEUTIC EXERCISES	8	0.5	267.6	268.1	12	0.3	142.0	142.3
76092	MAMMOGRAM, SCREENING	9	3.7	113.0	116.7	16	7.9	71.3	79.2
99203	OFFICE/OUTPATIENT VISIT, NEW	10	52.5	30.2	82.7	8	53.3	27.9	81.2
99245	OFFICE CONSULTATION	11	3.7	26.7	30.4	18	2.1	16.0	18.1
73721	MAGNETIC IMAGE, JOINT OF LEG	12	0.0	9.6	9.6	32	0.0	5.2	5.2
97530	THERAPEUTIC ACTIVITIES	13	0.0	120.8	120.8	63	0.0	47.2	47.2