



The BHCG/GNS Wisconsin Physician Value Study: Structural Issues Undermine its Effectiveness for Physician Referrals and Health System Assessments

EXECUTIVE SUMMARY

Initiatives focused on improving health care quality and efficiency are to be commended. Working both independently and collaboratively, health care providers, insurers and employers have made Wisconsin a national leader in health care quality, integrated health care delivery and the development of value-based payment programs. Efforts to provide meaningful and actionable information for health systems and employers can continue this legacy. Those efforts, however, must be based on accurate performance evaluations to prevent unnecessary disruption that causes more harm than good. They must also reflect Wisconsin's highly integrated health care systems, which achieve quality outcomes through care coordination.

Wisconsin's reputation for quality and innovation is due to the many commitments and investments that have been made by employers, health care providers, government agencies and insurers over the years. The state's integrated health systems have developed treatment protocols focused on best practices that improve patient outcomes and reduce the unnecessary duplication of testing and other medical services. By definition, integrated care should not be evaluated based on the performance of individual physicians. It should be assessed on the totality and consistency of medical care provided for as broad a spectrum of conditions as possible. Analyses limited to individual medical specialties fail to consider the successful efforts of integrated health delivery systems to improve the totality of the medical care they provide. In addition, they may inadvertently promote the fragmentation of care, which can lead to the duplication of tests, inconsistent care and a lack of coordination that would adversely impact outcomes. That could raise health plan costs and make employees less productive – the exact opposite of the results employers seek.

The recent report released by the Business Health Care Group (BHCG), the *Wisconsin Physician Value Study*, has the stated objectives of informing employer benefit program design and supporting performance improvements within health systems. It seeks to use the study's results to help employers steer patients to high-performing primary care physicians. Unfortunately, defining health system performance on a small subset of conditions treated by primary care physicians excludes the vast majority of medical care provided within integrated health systems. It also overstates the ability of primary care physicians to direct the care provided by specialists and other health care providers.

Even within its limited focus on primary care, the study was unable to confidently determine whether three-fourths of the physicians evaluated performed above or below the state average. That ambiguity makes it ineffective for referrals or performance evaluations: for every four physicians evaluated, three would be classified as “don’t know.”

There are several other issues that undermine the study’s intended objectives:

- Its one-year timeframe is too short for credible analysis; most analyses include three to five years of claims data
- Its focus is limited to 26 conditions representing just one-third of the episodes treated by primary care physicians¹
- It excludes 85 percent of the episodes and more than 90 percent of the medical utilization provided by health care systems²
- Its quality measures are focused on process measures rather than clinical outcomes, and excludes information that is not found in medical claims, including lab results
- It does not address the appropriateness of care, referral patterns and coordination of care within integrated health systems
- Its “black box” methodology makes it impossible for health systems to validate the credibility of its risk adjusters, which is critical to understanding whether the analysis is appropriately identifying top-performing physicians
- Its projected savings assumes that all of the state’s primary care could be moved to 261 physicians identified as “good” or “outstanding,” which is not feasible, or that primary care physicians will adapt best practices that the study does not identify

The BHCG study used commercial, Medicaid and Medicare fee-for-service claims data from the Wisconsin Health Information Organization (WHIO), a statewide, all-payer claims database that includes approximately two-thirds of the state’s population.³

WHIO’s data helps to provide direction and magnitude for potential efficiency improvements at the health system or provider group levels. Like satellite photos, it is useful for identifying relative positions and where things are headed, but it cannot “see beneath the trees.” Evaluating quality and providing granular analysis actionable at the individual provider level requires information that is not attainable through WHIO data, including lab results and other clinical data contained in electronic health records.

Using incomplete data to rank physicians, as is done in the BHCG study, is unlikely to improve health care delivery but could lead to unnecessary market disruption and undermine other, more credible efforts to identify best practices.

¹ The study identified primary care physicians as family practice physicians, internal medicine physicians and pediatricians.

² Because it is limited to 26 conditions, the analysis excludes 67% of episodes and 60% of medical utilization attributed to primary care physicians (Table 3). And because the study is limited to primary care physicians, it excludes 55% of total episodes and 78% of medical utilization (Table 4).

³ Based on WHIO DM release notes

THE BUSINESS HEALTH CARE GROUP ANALYSIS

BHCG, with funding from the Greater Milwaukee Business Foundation on Health, commissioned GNS Healthcare, a Boston-based analytics company, to review one year of claims data (January 1, 2017 to Dec. 31, 2017) from WHIO's all-payer claims database. The primary focus of the study was to identify the highest-quality/lowest-utilization primary care providers, which were defined as family practice physicians, internal medicine physicians and pediatricians.⁴ The study also sought to determine if there was a correlation between medical utilization and quality, as well as the savings that could be achieved if medical care could be moved to top-tier (high-quality, low-utilization) providers as defined by BHCG. The study included 456,753 patients, which represents an estimated 11 percent of the Wisconsin population.⁵ For the primary care physician analysis, GNS analyzed 26 conditions (Table 1).

TABLE 1. BASE ETGS INCLUDED IN THE PRIMARY CARE PHYSICIAN ANALYSIS⁶

Condition (Episode Treatment Group)	Percent of Physicians That Treat the Condition	Median Episodes Treated Per Physician	Episodes as % of Selected ETG Episodes
Hypertension	90%	100	23%
Diabetes	86%	43	10%
Mood Disorder, Depressed	96%	28	6%
Hyperlipidemia	82%	28	9%
Obesity	94%	25	8%
Asthma	98%	24	6%
Tonsillitis, Adenoiditis or Pharyngitis	95%	20	8%
Otitis	93%	14	6%
Acute Bronchitis	94%	11	4%
Acute Sinusitis	91%	10	4%
Ischemic Heart Disease	78%	8	3%
Chronic Obstructive Pulmonary Disease	81%	8	2%
Other Drug Dependency	85%	7	2%
Migraine Headaches	90%	6	2%
Mood Disorder, Bipolar	75%	3	1%
Cerebral Vascular Disease	74%	3	1%
Atrial Fibrillation & Flutter	71%	3	1%
Osteoporosis	67%	2	1%
Epilepsy	71%	2	1%
Congestive Heart Failure	70%	2	1%
Alcohol Dependence	68%	2	1%
Opioid Dependency	51%	1	1%
Inflammatory Bowel Disease	53%	1	<1%
Heart Failure, Diastolic	50%	1	<1%
Adult Rheumatoid Arthritis	54%	1	<1%
HIV/AIDS	6%	0	<1%

⁴ The study also included a very limited analysis of top-performing cardiologists, orthopedic surgeons and obstetricians. The specialist analysis was limited to one procedure for cardiologists (PTCA), two procedures for orthopedic surgeons (hip replacement and knee arthroplasty) and one procedure for obstetricians (pregnancy with delivery). Those results are not addressed directly in this white paper, but the general issues discussed (small sample sizes, confidence levels and lack of actionable results) would be of equal or greater concern.

⁵ Typically, 70% of an enrolled population have a medical claim in a given year. Given BHCG's assertion that 456,753 patients were included in the study, the total "enrolled" population would be 652,504, which represents 11% of the state's population in 2017

⁶ BSGA replicated the study's parameters using the same database to calculate the number of primary care physicians treating a given condition and the median number of episodes treated by each physician.

Quality ratings were based on how well a physician met Optum’s claim-based quality measures for a variety of treatment protocols.⁷ Efficiency was determined by comparing GNS’ proprietary risk-adjusted utilization for all services (inpatient, outpatient, professional, pharmacy, ancillary) for a provider against the statewide average for each of the 26 Base ETGs. A composite score for each physician was determined by aggregating the physician’s Base ETG performance scores using a weighted adjustment that reflected each physician’s episode mix for the Base ETGs.

The analysis included 3,760 primary care physicians who had at least 100 EBM opportunities in 2017. This represents approximately one-half of the total Wisconsin primary care physicians with a 2017 claim in the WHIO DM19 database.⁸ The study identified 141 primary care physicians statewide (4 percent of all physicians included in the analysis), as “outstanding” for both efficiency and quality at an 80-percent confidence level. Almost twice as many physicians were identified as “outstanding” for quality as were identified as “outstanding” for efficiency. Approximately one-third of providers were identified as having below-average quality and one-fifth were identified as having below-average efficiency (Table 2).

The study found no correlation between a provider’s quality score and his/her efficiency score. That is, a high-quality provider was as likely to outperform the statewide average in terms of efficiency as he/she was to underperform the average. This could be due to the study’s use of the Optum quality measures. While these measures may be helpful in identifying whether physicians are appropriately prescribing diagnostic tests and medications, they are not a valid tool for measuring improvements in a patient’s health status over time. Basing quality on process measures may also be the reason the study found no correlation between quality and efficiency: physicians were evaluated primarily on whether they ordered a test rather than their interpretation of the results and subsequent treatment plan.

Using an 80-percent confidence interval, the study identified 261 physicians who had an efficiency performance score that was above average (“outstanding” or “good” performers). An additional 784 physicians (21% of the total) were identified as having below-average performance scores. The study could not determine whether the remaining 2,715 performers – 72 percent of all the physicians analyzed – had above-average or below-average performance scores with 80-percent confidence.

The study estimated that medical service utilization could be reduced 29 percent if all primary care in the state was steered to the approximately 261 primary care providers who could be demonstrated to perform above the statewide average, or if all of the underperforming primary care providers changed their practice patterns to perform at an above-average level. Steering all Wisconsin patients seeking primary care to 261 physicians statewide is not feasible. Nor does the study provide any insights into the practices physicians could adopt to improve their performance scores.

TABLE 2. STRATIFICATION OF PRIMARY CARE PROVIDERS			
Physician Ranking	Quality	Efficiency	Both
Outstanding	502 physicians (13%)	260 physicians (7%)	141 physicians (4%)
Good	133 (4%)	1 (0%)	Not Provided
Typical	1,806 (48%)	2,715 (72%)	Not Provided
Below Average	1,319 (35%)	784 (21%)	Not Provided
BHCG Study Definitions:			
Outstanding: PCPs who perform better than the 75 th percentile score at an 80% confidence level			
Good: PCPs who perform better than 50 th percentile but not better than the 75 th percentile score at an 80% confidence level			
Typical: PCPs who could not be determined, with 80% confidence, to be better than the 75 th percentile score or below the 50 th percentile score			
Below Average: PCPs who perform below the 50 th percentile score at an 80% confidence level			

⁷ Optum’s EBM Connect product utilizes more than 600 claims-based measures to assess physicians’ compliance with treatment protocols; most of the measures assess whether a specific test or medication was prescribed.

⁸ Using the DM19 data mart, BSGA determined there were 7,457 unique primary care provider names with at least one claim in 2017

METHODOLOGICAL CONCERNS

BSGA identified six methodological concerns that mitigate the study's conclusions and impact.

1. The study's timeframe is too short

The BHCG analysis is limited to one year of medical claims data, which is significantly below the two, three or four years of claims data that most credible analyses require. The one-year window presents several challenges. It distorts data for physicians who onboarded new patients during the year. Several of the conditions included in the study are chronic diseases that require significantly more medical resources in the first year of care to mitigate health care utilization in subsequent years. The one-year window also distorts performance scores for physicians with patients who had an inpatient stay or catastrophic event within that year, much the same way a large claim can distort an employer's experience for a given year. A longer claims window would smooth the impact of catastrophic events and provide a more accurate depiction of a provider's performance over time.

The one-year window also limits the ability to credibly identify high- and low-performing providers. Because of its small sample size, the study was unable to identify whether 72 percent of the primary care providers had above-average or below-average efficiency at an 80-percent confidence level, which is much lower than the 90-percent or 95-percent confidence levels generally recognized as necessary to delineate statistically differences in performance.

As noted earlier, BHCG seeks to use the results of this study to steer patients to "high performing providers". However, the study could not indicate with at least 80 percent confidence that a significant number of primary care providers either outperformed or underperformed their peers. Steering patients away from these providers could not be supported based on the data, and it would result in large disruptions in care.

2. The study's focus is too narrow

By focusing on just 26 of the more than 445 Base ETGs attributed to primary care physicians, the analysis excludes two-thirds of the episodes and 60 percent of the medical resource utilization attributed to family practice physicians, internal medicine physicians and pediatricians who are included in the study (Table 3). These primary care physicians represent just half of the primary care physicians who had a claim in 2017 based on the WHIO data. Excluding 60 percent of medical utilization attributed to primary care physicians can distort results because it does not reflect the total range of services provided by primary care physicians. Additionally, the study excludes one-half of the state's primary care providers, especially new physicians who have received the most current training from medical schools.

CONDITIONS	EPISODES	% TOTAL	UTILIZATION	% TOTAL
26 conditions (Base ETGs) Included	780,528	33.3%	823,820,494	39.8%
Base ETGs Excluded	1,562,262	66.6%	1,242,959,170	60.2%
All Medical Care	2,342,790	100%	2,066,779,664	100%

The study also excludes nurse practitioners and physician assistants, two medical specialties that are providing an increasing share of primary medical care. In DM19, approximately 425,000 episodes for the selected Base ETGs were attributed to nurse practitioners and physician assistants in 2017. In addition, by limiting its focus to primary care physicians, the study excludes 55 percent of the episodes and nearly 80 percent of medical care provided to Wisconsin patients (Table 4).

⁹ Family Practice, Internal Medicine and Pediatric physicians with a completed episode in 2017 (WHO DM19)

TABLE 4. MEDICAL CARE UTILIZATION – ALL SPECIALTIES ¹⁰					
PHYSICIAN	EPISODES	% TOTAL		UTILIZATION	% TOTAL
Family Practice	2,119,425	26.7%		868,964,250	12.9%
Internal Medicine	753,216	9.5%		364,556,544	5.4%
Pediatricians	701,800	8.8%		235,130,000	3.5%
Included Care (Primary Care)	3,574,441	45.0%		1,468,623,794	21.8%
Excluded Care (Other Specialties)	4,361,597	55.0%		5,284,944,544	78.3%
All Medical Care	7,936,038	100%		6,753,568,338	100%

The study also ranked health systems based on the scores of their primary care physicians, although this portion of the study was not released publicly. BHCG has indicated it will share those results privately in its discussions with health systems.

Ranking health systems based on the performance score for primary care physicians raises several fundamental concerns. A credible systemwide score based on the aggregation of individual physician performance scores is not possible if the methodology cannot determine, for more than 70 percent of the physicians analyzed, whether the physician underperformed or outperformed their peers. In addition, basing a health system score on the performance of primary care physicians overestimates their ability to control the medical care provided outside of their offices. While some health plans have strict primary care gatekeepers, that is not true for many of the health plans serving the state. Many patients can see whichever specialists they choose. Even in strict gatekeeper models, primary care physicians have little control over the procedures performed by other physicians and health care professionals in inpatient and outpatient settings. Limiting health system scores to the performance of primary care physicians excludes the care provided by surgeons, oncologists, cardiologists, obstetricians, gynecologists, rheumatologists and other physician specialists; physician assistants; nurse practitioners and physical therapists. It also does not include any lab tests, diagnostic imaging or prescriptions ordered by these providers.

Measuring the totality of care provided to a patient within a health system provides a more accurate assessment of overall efficiency. For example, a primary care physician who is quick to refer patients to cardiologists for hypertension may appear to be more efficient than primary care physicians who continue to treat those patients even though the physician’s decision could result in less-efficient utilization of medical resources.

Finally, it has been a challenge for WHIO to properly attribute physicians to health systems.¹¹ In its analyses for various provider clients, BSGA has found that the number of providers inappropriately attributed to a health system can range between 10 and 20 percent of providers. This can even be greater for health systems that are expanding or consolidating.

3. The study uses “black box” technology that cannot be independently validated

Risk adjusting is critical to normalizing the medical utilization used to treat patients so that physicians are not penalized for having sicker patients. The risk-adjusting methodology used in most WHIO analyses relies on Optum’s severity adjustments, which are well documented and transparent.¹²

¹⁰ DM18 Commercial based on standard attribution. BSGA used DM18 to make these calculations because it is a significantly more robust database than DM19 and more representative of the medical care provided statewide. Two major carriers – Anthem, which primarily serves eastern Wisconsin and Health Traditions, which primarily serves western Wisconsin – did not participate in DM19

¹¹ Physician attribution was initially done by the Wisconsin Medical Society. In subsequent years, hospital systems voluntarily updated their attributions, with some health systems more diligent than others.

¹² Symmetry® Episode Treatment Groups:® Measuring Health Care with Meaningful Episodes of Care (Optum, Inc., 2017)

GNS Healthcare’s Reverse Engineering Forward Simulation (REFS) determined the relative weight of the study’s risk-adjusters based on its machine-learning algorithms. The risk-adjuster is critical to ensuring an accurate comparison of physician performance with respect to utilization when physicians are taking care of patients with the same diagnosis but different severities. Because the risk-adjustment mechanism has not been made available, it is impossible to assess the validity of the weighting assigned by GNS, whether the sample sizes for individual physicians are large enough to achieve credible weight adjustments and if the resulting physician rankings are accurate. In addition, if machine learning recalculates the risk-adjusting weights with every dataset analyzed, it could be challenging to conduct credible longitudinal analyses using datasets covering timespans.

4. The study’s quality measures focus on process measures, rather than outcomes

The Optum quality measures used in the analysis are only based on information that is included in claim forms. They can measure whether a blood test was performed, because there would be a billing code for that test, but they cannot measure the results of the blood test. They also do not include information about blood pressure, weight and other patient data recorded during office visits.

While the Optum quality measures may be helpful in identifying whether physicians are appropriately prescribing diagnostic tests and medications, they are not a good tool for measuring improvements in a patient’s health status over time. This could be the reason the study found no correlation between quality and efficiency: Quality was based primarily on whether a physician ordered a test rather than on his or her interpretation of the results and subsequent treatment plan.

There are additional challenges with Optum’s quality measures. They are not equally distributed among major practice categories, which would make it difficult to conduct a similar study for many specialties. Episodes typically treated by primary care physicians and cardiologists have the most quality measures, while pulmonologists, surgeons, orthopedic surgeons and others have relatively few. In addition, WHIO discontinued the publication of Optum’s quality measures in WHIO 2.0, its most recent release and has indicated they will not be included in future releases.¹³

Finally, the BHCG analysis does not identify any of the quality measures it used to determine quality scores. This makes it impossible to validate the appropriateness of the quality measures tied to each of the 26 conditions included in the analysis or to determine the areas where physician quality could be improved.

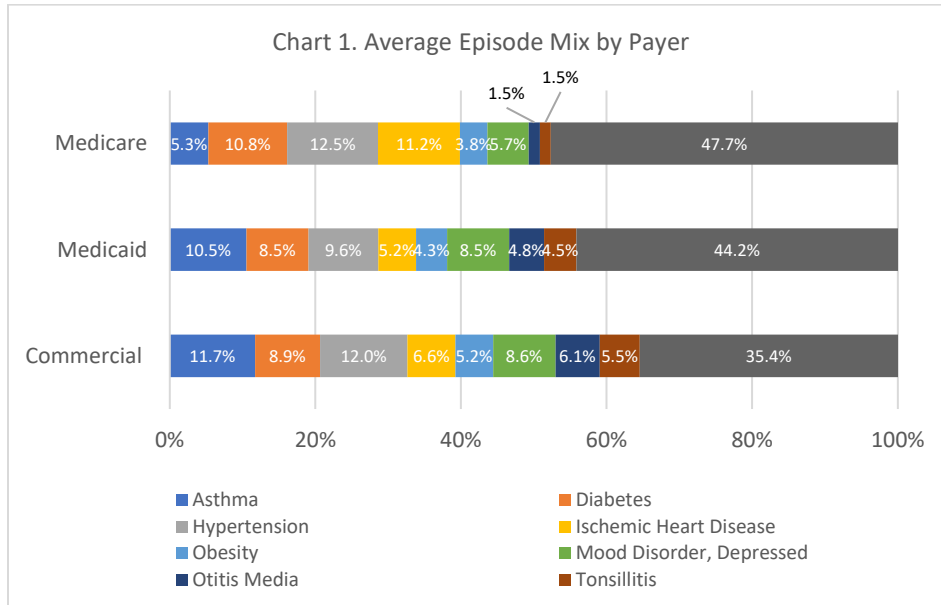
5. Comingling payers prohibits “apples-to-apples” comparisons

Variations in age and other demographics impact why patients seek medical treatments. Older patients are more likely to have chronic diseases such as osteoporosis, chronic obstructive pulmonary disease or ischemic heart disease. Younger patients are more likely to have tonsillitis or ear infections. Middle-aged patients have a higher prevalence of hypertension and obesity. Lower-income people may have more issues with alcohol dependence and depression. As a result, a physician’s episode mix reflects the relative proportion of commercial, Medicare and Medicaid patients they see. This is important because the methodology weights each physician’s performance based on his or her unique mix of episodes. Asthma will have twice the weight if physicians primarily treat commercial or Medicaid patients than if they treat Medicare patients. Conversely, ischemic heart disease is twice as important for providers who treat Medicare patients as it is for those who treat commercial or Medicaid patients (Chart 1).

In order to achieve more accurate comparisons at the individual provider level, payers should be segregated and/or normalized. A separate provider performance score should be determined for each of the three payer categories – commercial, Medicare and Medicaid – or every physician’s

¹³ WHIO 2.0 release notes and conversations with WHIO staff

episode mix should be normalized to common market basket of episodes that reflects a typical payer mix.



6. The expected savings in medical utilization are likely overstated

The WHIO database uses a “standard price”¹⁴ established by Optum, which provides the episode groupers. The standard price is designed to mitigate differences in negotiated discounts or the cost of living in geographic areas: \$1 in standard price in Milwaukee equals \$1 in standard price in Green Bay or La Crosse. This allows the standard price to be used as a proxy for utilization.

The “savings” cited in the BCHG study are identified as dollars, but are, in reality, a reflection of a reduction in utilization. Reducing inappropriate utilization can impact costs, but the impact will vary by provider and geographic area depending on the actual rates charged for individual services. Reducing inappropriate utilization in a low-cost area will have less financial impact than reducing inappropriate utilization in a high-cost area.

More importantly, the study’s inability to determine— at a minimal confidence level – whether 70 percent of physicians are outperforming or underperforming the state average for efficiency makes it impossible to identify which providers need improvement and why.

¹⁴ Also referred to as “standard cost”

LOOKING FORWARD

Primary care services constitute a small, but integral portion of the medical care provided to Wisconsin residents. In order to be credible, evaluations of primary care providers must measure care over a multi-year period, include a broad number of ETGs and use a transparent risk-adjustment methodology. Because primary care physicians have little control over the care provided outside of their offices, which is where most of medical utilization occurs, their efficiency scores should not be used to evaluate the totality of care provided in integrated health care delivery system. Integrated health systems should be evaluated on the totality of medical care provided by the full spectrum of health care professionals – primary care physicians, specialists, physician assistants, nurse practitioners, physical therapists and other professionals.

WHIO's all-payer claims database, when properly utilized, can benchmark health care performance and identify areas for improvement. However, WHIO data is insufficient to provide a complete evaluation of physician performance because it does not include clinical insights, outcomes information, lab results or electronic health records. Provider analyses should also incorporate, as much as possible, social determinants of health due to the impact that patient behavior and socioeconomic status has on health outcomes. Admittedly, there is very limited data in this area, but considerable work is being done to identify which determinants are most impactful.

As stated earlier, Wisconsin's reputation for quality and innovation is due to the many commitments and investments that have been made by employers, health care providers, government agencies and insurers over the years. The state's integrated health systems have developed treatment protocols focused on best practices that improve patient outcomes and reduce the unnecessary duplication of testing and other medical services. Analyses limited to individual medical specialties fail to consider the successful efforts of integrated health delivery systems to improve the totality of the medical care they provide. In addition, they may inadvertently promote the fragmentation of care, which can lead to the duplication of tests, inconsistent care and a lack of coordination that would adversely impact outcomes. That could raise health plan costs and make employees less productive – the exact opposite of the results employers seek.

As health care continues its transition to value-based payments, collaboration based on credible analyses and clinical insights, will foster continued improvement and properly align incentives to ensure the focus remains on quality, patient outcomes and efficient health care delivery.