The value of value-based care

Achieving Accurate, Transparent, and Replicable Monetary Estimates of Shared Savings in Value-Based Health Care

Aaron R. Wells, PhD, Vice President of Outcomes and Reporting, PopHealthCare, LLC, a GuideWell Company

Overview

With the passing of the Medicare Improvements for Patients & Providers Act in 2008, value-based healthcare has steadily become the cornerstone reimbursement methodology to change provider behavior from transaction-based care to longitudinal, patient-centered care. With the Centers for Medicare & Medicaid Services (CMS) Primary Cares Initiative launched in January 2020, CMS is aggressively moving the market toward value-based reimbursement. In the Commercial payer space, Anthem, Aetna, UnitedHealthcare Group, and Blue Cross Blue Shield of Florida have dedicated significant people, processes, and technology to create value-based arrangements across the healthcare landscape. All of these payers, representing the continuum of public and private insurers, are striving toward a common aim initiated by CMS – reform how health care is delivered and paid for in order to better care for individuals, better health for populations, and lower cost by moving toward paying providers based on the quality, rather than the quantity, of care provided to patients.[1]

Over the last 12 years, myriad methods have been tested for monetizing the "value" in value-based care, with no clear indication regarding the methodology yielding the highest fidelity estimate. Underscoring this point is the current use of two different methodologies by the Center for Medicare and Medicaid Innovation (CMMI) for monetizing value derived by the Next Generation Accountable Care Organization (NGACO) Pilot.[2] Further complicating ambiguity in monetizing value-based care is the lack of standards or best practices. This article aims to provide an overview of monetizing value-based healthcare and highlight Coarsened Exact Matching as the prominent method for monetizing value; a follow-up article will enumerate a set of best practices providers and payers can follow when entering into value-based contracts.

Monetizing Value-Based Healthcare

In healthcare contracting, the term "value" denotes the monetary equivalent of a system of connected, timely, and appropriate care for a given individual that results in a better health outcome than was anticipated. Value, in this context, is a macro level measure of both what does occur (e.g., coordinate care across multiple providers) and does not occur (e.g., unplanned acute hospitalization). Specific to

the latter measure of value, specialized methods are required to monetize health care externalities. Principle methods utilized by prominent leaders in value-based contracting include quasi-experimental and actuarial trend-based methods. Quasi-experimental methods are predicated on retrospective evaluation of a common outcome observed in at least two cohorts (treated, untreated) from the same population matched on a shared set of individual-specific attributes over a minimum two time period basis. Actuarial trend-based methods vary in methodological specificity but generally rely on two or more years of longitudinal data for the population under measurement combined with factors related to technology, income, inflation, and morbidity to derive an empirical estimate of the expected trend sans the health improvement program.[3]

While methodologically distinct, both quasi-experimental and actuarial trend-based methods rely on the presence of data to estimate the monetary value of absent data. In other words, if a given value-based program is effective, patients enrolled in the program will have fewer billable health care events during the evaluation window (absent data) than a similar group of patients not enrolled in the program (present data). Using as an example one of the earliest value-based programs – Hospital Readmissions Reduction Program – providers were no longer to be paid for readmissions of the same diagnosis occurring within 30 days of the index admission (generalized framing). Thus, providers who did not generate "readmission data" remained financially neutral for their admitted patients. In another example, the NGACO program, participating providers could earn shared savings – a monetary equivalent of value – by keeping their attributed patient population at or below specified adverse thresholds as well as at or above certain positive thresholds. The adverse thresholds principally centered on an estimate of expected per beneficiary paid medical and pharmacy claims. NGACO participants who successfully pursued strategies to reduce unwarranted treatment variation, connect patient providers, support post-discharge care, focus on near- and long-term individual health outcomes, increase patient and provider satisfaction, and overall, create a more sustainable, higher quality healthcare delivery network were rewarded. The reward was a monetary payment from CMS to the NGACO participant quantified as a function of the difference between the NGACO's observed Measurement period performance and the expected performance level based on the participant's peer group.

Converting Value into Shared Savings via the Coarsened Exact Matching Methodology

Referenced earlier, two primary approaches to monetizing shared savings in a value-based contract are actuarial and quasi-experimental. The purpose of this document is to focus on the latter, but it is important – given the historical, current, and broad application of actuarial methods – to provide a limited overview of these methods. Given extensive variability in actuarial methods, one class of methods most relevant here is the class utilized by CMMI for NGACO financial performance evaluation. Specifically, CMMI monetizes value created by an NGACO by comparing the entity's actual total cost of care among attributed beneficiaries to a prospectively computed benchmark for the same performance year, where the benchmark is a forecast based on historical performance of the ACO's beneficiaries.[4] The methodology for computing the forecast is detailed in the NGACO benchmarking methods documents.[5] The fundamental flaw of this class of actuarial methods is unfortunately the very fact that is promoted as the core feature of the methodology, specifically, that neither the baseline expenditure data nor the projected regional trend will be updated after calculation of the benchmark.[6] The most illustrative demonstration of this fundamental flaw is the transformative effect coronavirus strain COVID-19 has had on health care utilization during 2020 in which utilization has declined precipitously and unexpectedly in compliance with quarantine laws, individual trepidation to seek care, and furloughed providers.[7] Due to the fixed benchmark, CMMI and other payers who have contracted with such a methodology will be paying out significantly larger sums of shared savings.[8] This author believes CMMI and payers using the class of actuarial methods based on forecasts and fixed benchmarks will

jettison this methodology in favor of quasi-experimental methods in 2021 forward in order to account for exogenous, unexpected effects on health care trend. In other words, COVID-19 has shown payers that the past is no longer a reliable indicator of the future when the future becomes the present.

It is worthwhile to note that in this author's experience, some clients prefer to asses financial performance of vendor provided value-based services by simply comparing the trend in per member per month total cost of care between the enrolled and program eligible yet not enrolled members period over period. This approach, which incorporates the trend normalization property of a difference-in-difference method and controls at a macro level for intra-cohort member differences (i.e., selection bias), relies on the erroneous assumption that being eligible for the program removes selection bias. In contractual settings where this simplistic method is employed, which is typically championed by non-actuarial individuals, both parties to the contract must show that use of a quasi-experimental valuation methodology that accounts for selection bias yields a statistically similar estimate. If results of the two different methods yield statistically significantly different results, the quasi-experimental estimate should be relied upon as it directly accounts for selection bias in the monetization of program value.

One such prominent quasi-experimental method for value-based healthcare applications is Coarsened Exact Matching (CEM).[9] Co-developed by Gary King, Weatherhead University Professor at Harvard University and Director of the Institute for Quantitative Social Science, CEM is a non-parametric method that is relatively simple to implement, transparent, and easily understood by non-quantitative stakeholders. CEM originated in the political science discipline; the earliest application of CEM in a US health care setting was conducted by Wells et al. (2013)[10] and recently, CEM was applied to evaluate the effectiveness of nurse practitioner dementia care co-management on acute care utilization, long-term care admissions, and hospice use outcomes.[11] In contrast, the quasi-experimental method Propensity Score Matching (PSM) has been utilized in US healthcare applications for decades and is currently employed by CMMI for evaluating the NGACO pilot. Despite the tenured use of PSM, the method is complex, assumption based, not readily replicable, error and bias prone, and inconsistent on a theoretical basis with value-based care monetization.

Underscoring the limitations of PSM, Harris and Horst (2016)[12] summarized the process of implementing an application of PSM to six steps, with each step involving a series of decisions by an analyst. These decisions are guided by both objective, empirical evidence as well as subjective input. The existence of the latter compounds the complexity of PSM, ultimately resulting in this method conforming to the setting of each new application instead of providing a rigorous, standardized methodology for monetizing value. Recently, King et al. (2019) showed the empirical ramifications of PSM for estimating causal effects and concluded the method can yield matched populations with more imbalance than matches created from random sampling.[13]

Use of CEM requires only two key decisions by the parties to a value-based contract – determination of the matching factors and binning levels for non-binary factors. Even these two steps can be automated, though, in this author's experience, the available options for automated matching factor selection and binning have not been sufficiently tuned to the non-normal distribution of healthcare factors. Once the matching factors (F) and bins (B) have been selected, the combination of which creates $F \times B$ strata into which members are uniquely assigned to only one stratum, estimation of the program effect within CEM simply involves four steps. First, compute the mean of the outcome by strata, cohort, and time period; in most applications, there are only two cohorts (treated, untreated) and time periods (pre-program or baseline and intervention or measurement period). Second, by strata and within each cohort, subtract the intervention period outcome mean from the baseline outcome mean (yields first difference). Third,

exactly match the cohorts by strata and subtract the treated from untreated cohort first difference values (yields second difference). Steps two and three result in the marginal or strata-level difference-in-difference (DID) value of the outcome; this DID value is a detrended measure of the outcome controlling directly for observable member-level attributes. The last step involves three simple, linear mathematic operations: (a) multiply the strata level DID value by the weight specific to the strata, which in turn is a linear mathematical function of the cohort by strata distribution; (b) sum the weighted DID values over all strata; and (c) divide (b) by the total number of intervention period treated weights. The resulting value from Step (c) is the estimated treatment effect of the value-based program.

The construction of CEM is such that an end-user can directly connect value loss or gain between a given treated individual and the health intervention program under investigation. The term "directly" means that no transformation of the individual's attributes or outcome measure occurred and more importantly, the intervened individual's comparison cohort is identifiable due to exactly matching on observable features. In contrast, PSM relies on a statistical model with estimation inaccuracy, distributional assumptions, independence of irrelevant alternatives assumption, calipers for pruning observations, algorithm choice for assigning untreated members to treated members, and lastly, the ratio of untreated to treated member matches – all of which results in opaque insights into the connection of value to the health intervention program. CEM enables a stakeholder to easily identify specific subsets of the intervened population and investigate the attributes associated with this subset, their matched non-intervened peers, and the outcomes under study. For stakeholders, the ease and transparency of CEM-based results means they can pinpoint issues with the health intervention program and develop protocols to address the issues quickly.

In addition to CEM allowing for unambiguous attribution of value, the methodology aligns with the concept of a provider managing each patient within a panel of similar patients. For example, a provider will follow a set of protocols for the management of her male patients having diabetes, congestive heart failure, mild depression, and advanced age. However, the provider will tailor her treatment regime to the unique set of each patient's attributes, as the protocols provide general medical treatment guidance but are not individualized. PSM, though, does not align with this theoretical similarity by matching patients at the individual level, or in some specifications, to a user-defined, globally applied ratio of untreated to treated patients. In summary, entities entering into value-based agreements should be aware of the unrealistic theoretical and empirical foundation of PSM and consider as an alternative CEM.

Conclusion

This paper provided an overview of a significant issue with the transition from fee-for-service to value-based healthcare – how to monetize what did not occur for the purpose of determining shared savings or penalties within a contractual arrangement between a payor and provider. To address empirically the inherent problems associated with monetizing healthcare utilization that does not occur, this paper presented the innovative methodology Coarsened Exact Matching and contrasted this method with other current methods. For providers, use of a robust methodology for computing the value created across their panel for employers, heath plans, and self-insured patients increases the likelihood of higher reimbursement, positive exposure as a high-quality provider, and direct insight into the practices responsible for the positive outcomes. A supporting example of this potential result is the NGACO pilot in which participating providers could have been allocated 107% more shared savings if a methodology similar to CEM had been applied rather than the actuarial method. For health plans, use of a robust methodology standardizes monetization across a diverse set of contracted entities, which creates labor,

technology, legal, and policy efficiencies while providing transparent and direct insights into program performance. Moreover, attributing program performance across multiple programs members are engaged in allows the plan to ensure duplication of value does not occur, and, concomitantly, value is accurately assigned to the contracted entity. Lastly, and most importantly, for patients who are likely to never know how the care they receive translates into shared savings or penalties for their providers, the benefits of value-based care will be realized by improved health outcomes, lower out of pocket health expenditures, and more time living outside of the healthcare system.

Aaron R. Wells, PhD is the Vice President of Outcomes and Reporting at PopHealthCare, a licensed provider group that partners with health plans to provide risk adjustment services and in-home medical care to high-risk health plan members in 15 states. Aaron is an accomplished and high-performing leader with expertise in value-based healthcare outcomes monetization, machine learning, and quasi experimental outcomes research. Aaron has extensive knowledge of and experience with statistical software applications and various types of health care data sets to qualitatively and quantitatively understand current and future trends and outcomes.

<u>References</u>

See https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs.

Refer to https://innovation.cms.gov/innovation-models/next-generation-aco-model for more information.

See as an example: Getzen, Thomas A. 2019. Getzen Model of Long-Run Medical Cost Trends: Technical Manual. Website Copy at https://www.soa.org/globalassets/assets/files/research/research-2016-getzen-model-tech-manual-doc.pdf.

Seema Verma. "Number Of ACOs Taking Downside Risk Doubles Under 'Pathways To Success.'" Health Affairs Blog, January 10, 2020. Available at https://www.healthaffairs.org/do/10.1377/hblog20200110.9101/full/

See https://innovation.cms.gov/files/x/nextgenaco-methodology.pdf and https://innovation.cms.gov/files/x/nextgenaco-benchmarkmethodology-py4.pdf.

See page 1 of https://innovation.cms.gov/files/x/nextgenaco-methodology.pdf.

See Robin Gelburd. 2020. Health Care Professionals and the Impact of COVID-19: A Comparative Study of Revenue and Utilization. American Journal of Managed Care. Available at https://www.ajmc.com/view/health-care-professionals-and-the-impact-of-covid-19-a-comparative-study-of-revenue-and-utilization.

A recent rule issued by the Centers for Medicare & Medicaid Services states that CMS will be smoothing over the COVID-19 reduced utilization period using a modified version of their

forecasting model. Accordingly, the benchmark is not staying fixed and instead adjusted to account for COVID-19. See rule https://www.federalregister.gov/documents/2020/05/08/2020-09608/medicare-and-medicaid-programs-basic-health-program-and-exchanges-additional-policy-and-regulatory#h-54.

Stefano M. Iacus, Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." Political Analysis, 20, 1, Pp. 1--24. Website Copy at https://j.mp/2nRpUHQ

Wells, Aaron R, Hamar, Brent, Bradley, Chastity, Gandy, William M, Harrison, Patricia L, Sidney, James A, Coberley, Carter R, Rula, Elizabeth Y, Pope,James E. 2013. "Exploring Robust Methods for Evaluating Treatment and Comparison Groups in Chronic Care Management Programs".Population Health Management. 16(1):35-45. Available at: http://info.healthways.com/hubfs/Science_and_Research/Exploring-Robust-Methods.pdf?t=1499459540214

Jennings LA, Hollands S, Keeler E, Wenger NS, Reuben DB. The Effects of Dementia Care Co-Management on Acute Care, Hospice, and Long-Term Care Utilization [published online ahead of print, 2020 Jun 23]. *J Am Geriatr Soc*. 2020;10.1111/jgs.16667. doi:10.1111/jgs.16667

Harris, Heather and Horst, S. Jeanne (2016) "A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process," Practical Assessment, Research, and Evaluation: Vol. 21 , Article 4. DOI: https://doi.org/10.7275/yq7r-4820. Available at: https://scholarworks.umass.edu/pare/vol21/iss1/4

Gary King and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching." Political Analysis, 27, 4. Copy at https://j.mp/2ovYGsW