Supplementary Materials

Included materials:

1.   Detailed Description of Quantitative Methods

2.   Appendix Table 1: List of measures and weights

3.   Appendix Table 2: Attribute scoring grid

Detailed Description of Quantitative Methods

Data Sources

QuintilesIMS' PharMetrics Plus[TM] Data is a proprietary database of U.S. commercial medical and pharmacy claims data that covers 150 million covered lives from 2006. Commercial insurance data reflects market prices, rather than prices set by Medicare and Medicaid, and allows analysis of the "all-in" cost of care for patients, including payments for drugs, ER visits, hospitalizations, lab testing and other services. The PharMetrics Plus[TM] data is derived from health insurance plans across the US and represents a diverse mix of commercially insured patients. Approximately 71% of the patients in PharMetrics Plus[TM] are covered by a PPO plan. There is limited inclusion of Medicaid patients. Data for eligible enrollees was extracted from PharMetrics Plus[TM]. Eligible enrollees included only those aged 0-64 with 12 months enrollment during the year, with prescription benefit for the duration of enrollment. Additionally, allowed amounts in the study year had to fall in the acceptable range of $10-$1,000,000. Our primary care study made use of PharMetrics Plus[TM] data from January 2009 – December 2011.

This claims data was combined with QuintilesIMS' OneKey[TM] Data (formerly IMS Health Care Relational Data), a proprietary database which provides comprehensive demographic information, address intelligence, affiliations and ownership relationships for over 4.4 million

professionals and 500,000 health care facilities. The OneKey^{TM} data allowed for meaningful specialty designation and aggregation of providers at the provider group level.

Providers included for analysis belong to the primary care-related specialties of internal medicine, family practice, geriatrics, pediatrics, and OB/GYN. Pediatricians and OB/GYNs were included in the analysis because they can serve as PCPs during childhood and child-bearing years, periods of particular relevance to a commercially-insured, under 65 population. In addition, many family medicine physicians deliver babies and care for children.

OneKey^{TM} was used to identify practice sites (e.g., physicians practicing together at a single site) from these specialties. Practice sites were defined as physicians affiliated and working with a defined outpatient medical practice, including multi-professional practices and sole physician practices. All cost and quality composite calculations described below took place at the practice site level.

Measurement guidance panel

We convened a scientific panel to provide guidance on our methodology. The panel included John Adams, PhD, Bill Thomas, PhD and Mark Rattray, MD for the cost methodology, and Ashish Jha, MD and Patrick Romano, MD for the quality methodology. The panel engaged in several conversations before, during, and after analysis of the quantitative data to arrive at consensus for our approach to attribution, risk adjustment, and outlier trimming on the cost side, and to measure selection and weighting and composite approach on the quality side.

Cost measures

We took a total cost of care perspective, including all medical and prescription claims aggregated for a patient during each 12-month period in which they were eligible. The unit of observation of costs was the patient-year. Because we used three years of aggregated data, a single patient could appear as up to three observations.

All cost analytics described below were conducted for two types of cost: payor allowed (i.e., reflecting negotiated prices) and standardized (i.e., assessed against a fee schedule). The fee schedule for standardized cost was developed based on the average payor allowed cost for each service across the entire PharMetrics Plus^TM Dataset. Variation in standardized cost reflects differences in utilization and intensity of health care services, while payor allowed cost reflects those differences as well as differences driven by contractual arrangements.

Attribution for cost analysis

Medical claims for enrollees eligible for attribution were collected if the rendering provider fell into one of the PCP categories. Each selected enrollee was attributed to the individual PCP with the highest number of claims (and then to that provider's practice if applicable). In case of ties, the following were applied as tie-breakers, in order of priority: (1) Latest date of service, (2) Earliest date of service, (3) Total allowed amount, and (4) Group before Individual.

Risk adjustment and outlier trimming for cost analysis

Each patient-year observation was assigned to a disease category using 3M's Clinical Risk Group (CRG) software to form cost-homogeneous groups of patients.[1-3] CRGs were included

only if a minimum of 150 enrollees across the full sample fall into the category. For each provider group, the PPPM cost was calculated for each CRG category. Cost outlier trimming was performed within each CRG (and pharmacy benefit status if deemed appropriate) in order to eliminate extremely low or high costs. Outlier thresholds were determined using the LogMean methodology for each CRG where total cost was trimmed for selected enrollees if total cost was outside the range of trim values for the enrollee's CRG. This process yielded an observed PPPM by CRG for each attributed medical group.

The observed/expected (O/E) PPPM ratio by CRG for each practice site was then calculated. The expected PPPM for a practice site was calculated at the level of the CRG and is the average PPPM across all qualified patients in that CRG. For each site, the average O/E ratio (Relative Cost Index) was calculated as the patient-volume-weighted average of each CRG-O/E PPPM ratio for each CRG for which the group has at least 30 enrollees attributed (after trimming). Put differently, a practice site's O/E ratio in a particular CRG was more heavily weighted in that practice site's Relative Cost Index if that CRG represented a greater share of the practice site's patient population. Finally, for each qualifying practice site the percentile rank and confidence interval were estimated around the site's Relative Cost Index.

Sources of cost difference were investigated by assigning claims to type and place of service by CRG. Inpatient stays were grouped using APR-DRGs and outpatient surgeries were grouped using a QuintilesIMS proprietary grouper. Remaining claims were categorized based on CPT, Revenue and NDC codes. PPPM spend by category was standardized using the spending by type of service for the CRG to which the patient was assigned.

Quality measures

We developed a composite measure, drawing 65 measures from the library provided by QuintilesIMS. We selected measures that use administrative claims data and are either in common use (through HEDIS or a CMS program such as Medicare STAR) or endorsed by the National Quality Forum. Four measures were excluded either because the expert clinical felt they lacked sufficient clinical significance for inclusion or were duplicative and 20 were excluded because they did not meet the sample size requirements described below. The final composite included 41 measures across the 5 broad domains listed below.

- Medication management compliance
- Medication management prescribing quality
- Medication management monitoring
- Prevention/wellness
- Treatment process of care

The full measure list is provided in Appendix Table 1.

Attribution for quality measures

Attribution for quality measures was based on rendering provider. An 'encounter table' was created containing provider, patient, and date of encounter, where provider belonged to the specialties considered PCP, encounters were identified by CPT codes for outpatient encounters, and dates varied by measure. The encounter dates covered the timeframe for eligibility (denominator) and numerator. To be eligible for attribution, a provider needed 2 or more

encounters. A single provider was chosen who had the most frequent number of encounters over the timeframe for that measure. In the case of ties, the provider having the more recent encounter was chosen. Attributed providers were aggregated for comparison at the practice level.

Quality rating calculation, risk adjustment and exclusions

The provider rating was based on the indirectly standardized measure of observed to expected for each measure. The Indirectly Standardized Composite (ISC) methodology produced a ratio of observed and expected numerators across the individual component measures, which were weighted according to the specified composite weights.

For a particular measure, the observed numerator was simply the number of times that a provider met the specified quality standard. The expected numerator was calculated as the peer group's rate of meeting quality opportunities for that measure multiplied by the provider's total number of quality opportunities for that measure. The peer group included all attributed medical groups for a particular measure. The ISC methodology used the ratio of the weighted sum of observed numerators to the weighted sum of expected numerators.

For a composite measure to be calculated and scored for a practice site, that practice site and its peer group satisfied the sample size requirements for calculating those measures, as follows:

- Required number of denominator for the practice site in the *composite*: The sum of the site's denominators across all individual measures that comprise the site's composite must be at least 30 in order for the practice site to be scored.

- Required number of peer denominator by *component*: The sum of component denominators across all practice sites in the peer group must be at least 30 in order for a peer rate for the component measure to be calculated. All attributed patients, regardless of whether the responsible site qualifies for scoring, are included in the calculation of peer rates.

- Required Number of Measures: A practice site can only get a composite score if they have a valid rate for at least 4 component measures, at least 1 of which must be from each of following two domains: medication management compliance and medication management prescribing quality.

A particular site's composite was only comprised of those measures for which the site had adequate sample size to produce a score.

Quality Measure Weights

A composite weight method was used in order to combine an empirical perspective (to weight those measures which most differentiated between providers) and a clinical perspective (to account for which measures had the most clinical significance and were under greater control of providers). The full measure list, along with weights, is provided in Appendix Table 1.

Empirical weights:

Empirical weights were created using the following method:

1. Exploratory factor analysis was used on the quality measures to determine the number of identifiable factors. Using a Scree plot, the expert panel decided on a cutoff of 12 factors. The minimum Eigenvalue of these 12 factors was 1.285.

2. Principal components factor analysis (constraining to 12 factors) was then used to identify the 41 measures' individual loadings onto each of the 12 factors.

3. These factors loadings were converted to measure-specific empirical weights using a modified calculation of each measure's communality (sum of square of factor loadings). The empirical weights used were the weighted sum of square of factor loadings, where the weights used were the Eigenvalues of each particular factor. In this way, the empirical weights account for both the relative importance of the factor (through the Eigenvalue weight) and the measure's loading onto that factor (square of factor loading).

Clinical weights:

The expert panel engaged in several conversations to arrive at consensus clinical weights for each of the 41 measures that were based on the panelist's clinical knowledge and expertise in quality measurement. In particular, greater weight was assigned to measures under greater control of the primary care provider (those domains covering medication management compliance and medication management prescribing quality). In contrast, measures belonging to three domains were given comparatively less weight (medication management monitoring, prevention and wellness, and treatment process of care).

Composite Weight:

A composite weight was derived for each individual measure as the product of the two weights (empirical and clinical), normalized so that the sum of weights is equal to 1.

Exclusion of Practices with only Pediatricians and OB/GYNs

Practices with only pediatricians and OB/GYNs were not included in purposeful sampling because their patient mix and quality measure representation were very different than that of the other types of practices. A post hoc sensitivity analysis was run to confirm that the practices visited would have ranked in the same value cohorts had we excluded these practices from the ranking.

Appendix References

1.   Hughes JS, Averill RF, Eisenhandler J, et al. Clinical Risk Groups (CRGs): a classification system for risk-adjusted capitation-based payment and health care management. *Med Care.* 2004;42(1):81-90.
2.   Johnson TL, Brewer D, Estacio R, et al. Augmenting Predictive Modeling Tools with Clinical Insights for Care Coordination Program Design and Implementation. *EGEMS (Wash DC).* 2015;3(1):1181.
3.   Pfister DG, Rubin DM, Elkin EB, et al. Risk Adjusting Survival Outcomes in Hospitals That Treat Patients With Cancer Without Information on Cancer Stage. *JAMA Oncol.* 2015;1(9):1303-1310.

Appendix Table 1. List of Measures and weights

Domain	Measure Name (HEDIS 2013 unless starred)	Clinical Weights	Empirical Weights	Composite Weights	Normalized Composite Weight
A	B	C	D	E=C x D	F = E / Sum(E)
Med Mgmt Compliance	Proportion of Days Covered Diabetes Roll-up*	0.3755	0.6472	0.2430	0.0393
Med Mgmt Compliance	Proportion of Days Covered Renin Angiotensin System Antagonists*	0.3755	0.7693	0.2889	0.0468
Med Mgmt Compliance	Proportion of Days Covered Statin*	0.3755	0.7582	0.2847	0.0461
Med Mgmt Prescribing Quality	Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis	0.2253	0.0650	0.0146	0.0024
Med Mgmt Prescribing Quality	Asthma Medication Ratio	0.2253	1.1355	0.2559	0.0414
Med Mgmt Prescribing Quality	Disease-Modifying Anti-Rheumatic Drug Therapy for Rheumatoid Arthritis	0.2253	0.2001	0.0451	0.0073
Med Mgmt Prescribing Quality	Use of Appropriate Medications for People with Asthma	0.2253	1.1723	0.2641	0.0428
Med Mgmt Prescribing Quality	Diabetes – Appropriate Treatment of Hypertension*	0.2253	0.6182	0.1393	0.0226
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: ACE or ARB	0.1288	1.7457	0.2248	0.0364
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Digoxin	0.1288	0.8929	0.1150	0.0186
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Diuretics	0.1288	1.6171	0.2082	0.0337
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Phenobarbital	0.1288	1.0980	0.1414	0.0229
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Phenytoin	0.1288	0.8494	0.1094	0.0177
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Miscellaneous Anticonvulsants	0.1288	0.3541	0.0456	0.0074
Medication Mgmt Monitoring	Annual Monitoring for Patients on Persistent Medications: Carbamazepine	0.1288	0.3292	0.0424	0.0069
Prevention-wellness	Adult BMI Assessment	0.1352	0.5595	0.0756	0.0122
Prevention-wellness	Breast Cancer Screening	0.1352	0.8790	0.1188	0.0192
Prevention-wellness	Cervical Cancer Screening	0.1352	0.9631	0.1302	0.0211

Domain	Measure Name (HEDIS 2013 unless starred)	Clinical Weights	Empirical Weights	Composite Weights	Normalized Composite Weight
A	B	C	D	E=C x D	F = E / Sum(E)
Prevention-wellness	Chlamydia Screening in Women	0.1352	0.2780	0.0376	0.0061
Prevention-wellness	Colorectal Cancer Screening	0.1352	0.4114	0.0556	0.0090
Prevention-wellness	Weight Assessment and Counseling for Nutrition and Physical Activity: Assessment of BMI Percentile	0.1352	0.9512	0.1286	0.0208
Prevention-wellness	Weight Assessment and Counseling for Nutrition and Physical Activity: Outpatient Visit/Nutrition Conseling	0.1352	0.8704	0.1177	0.0191
Prevention-wellness	Weight Assessment and Counseling for Nutrition and Physical Activity: Outpatient Visit/Physical Activity Counseling	0.1352	0.9741	0.1317	0.0213
Treatment Process of Care	Antidepressant Medication Management: Effective Acute Phase Treatment	0.3755	1.3490	0.5066	0.0820
Treatment Process of Care	Antidepressant Medication Management: Effective Continuation Phase Treatment	0.3755	1.3707	0.5147	0.0833
Treatment Process of Care	Comprehensive Diabetes Care: Hemoglobin A1c Testing	0.1352	1.5108	0.2042	0.0331
Treatment Process of Care	Comprehensive Diabetes Care: Eye exam (retinal) Performed	0.1352	0.3400	0.0460	0.0074
Treatment Process of Care	Comprehensive Diabetes Care: LDL-C Screening	0.1352	1.6050	0.2170	0.0351
Treatment Process of Care	Comprehensive Diabetes Care: Medical Attention for Nephropathy	0.1352	1.0473	0.1416	0.0229
Treatment Process of Care	Cholesterol Management for Patients with Cardiovascular Conditions	0.1352	0.8089	0.1094	0.0177
Treatment Process of Care	Appropriate Testing for Children with Pharyngitis	0.1352	0.6102	0.0825	0.0134
Treatment Process of Care	Follow-Up After Hospitalization for Mental Illness: 30 Day Follow-up	0.1352	1.3065	0.1766	0.0286
Treatment Process of Care	Follow-Up After Hospitalization for Mental Illness: 7 Day Follow-up	0.1352	1.2944	0.1750	0.0283
Treatment Process of Care	Use of Imaging Studies for Low Back Pain	0.1352	0.0920	0.0124	0.0020
Treatment Process of Care	Pharmacotherapy of COPD Exacerbation: Systemic Corticosteroid Dispensed Within 14 Days of Event	0.2253	0.9757	0.2199	0.0356
				Composite	Normalized

Domain	Measure Name (HEDIS 2013 unless starred)	Clinical Weights	Empirical Weights	Weights	Composite Weight
A	B	C	D	E=C x D	F = E / Sum(E)
Treatment Process of Care	Pharmacotherapy of COPD Exacerbation: Bronchodilator Dispensed Within 30 Days of Event	0.2253	0.9868	0.2224	0.0360
Treatment Process of Care	Use of Spirometry Testing in the Assessment and Diagnosis of COPD	0.1352	0.1797	0.0243	0.0039
Treatment Process of Care	Appropriate Treatment for Children with Upper Respiratory Infection	0.2253	0.6856	0.1545	0.0250

*Measure sourced from the Pharmacy Quality Alliance and used in CMS's Medicare Stars program at the time of the analysis

Appendix Table 2. Attribute scoring grid

Ordinally scored	Score 1	Score 3	Score 5
1. Expanded Access	Patients are frequently unable to get a same day appointment with their primary care provider or anyone else at the practice. There is no access before 9am or after 5pm and when the practice is closed, the answering message directs patients to the emergency department.	Some slots are left open for same day access, though most appointments are booked in advance. There are either extended hours beyond 9AM-5PM or a means of accessing a clinician when the practice is closed. This could include a nurse phone line or diversion to an urgent care center.	Patients call the practice and are offered a prompt appointment, ideally on the same or next day, no matter what the reason for the visit. Physicians keep a percentage of their appointments open to accommodate this. No patient is ever turned away. The clinic provides appointments beyond the hours of 8AM - 5 PM M-F and/or patients know they can reach a member of the care team on the phone after hours.
2. Clinician decision support for evidence-based medicine	There is little to no discussion between clinicians of common approaches to patient care. There is no evidence of the practice maintaining schedules of needed follow-up care for chronic illness or preventative services, doing outreach to proactively close care gaps or flagging this at the patient visit.	Clinicians treat patients similarly and talk about sharing a common practice style. There are flags in place at the time of the patient visit for evidence-based care and preventive services, however there is no outreach to bring patients in proactively and following up on care happens inconsistently. There is some evidence of providers only ordering clinically indicated tests and treatments for their patients, but this is inconsistent and/or not driven by evidence-based guidance.	The clinicians have established consensus and developed guideline-based reminders in the form of prompts made available at the time of the patient visit (i.e., in an EMR or structured encounter form). Routine guideline-based reminders are sent to patients and outreach takes place to ensure care gaps are closed for recommended preventive care. Physicians only order diagnostic tests if the result will likely influence their treatment plan.

Ordinally scored	Score 1	Score 3	Score 5
3. Risk-stratified care management	There is no risk stratification in place and no extra support for patients with greater needs.	There is some evidence of risk stratification and case management for patients with greater needs.	Provider skill sets are matched to patient needs. Systems are in place for identifying high-risk patients and the practice uses specially trained care managers whose primary job is to coordinate and improve the quality of care for patients with greater needs.
4. Shared decision-making and advanced care planning	Providers are likely to default to ordering more tests and procedures when managing patients' care and would never talk a patient out of a requested test or treatment likely to be of limited benefit. Physicians do not have the time or capacity to have conversations with patients when there are multiple diagnostic and treatment options. They either tell the patient what to do or leave it up to the patient without explaining the tradeoffs. There is no documentation of end of life care wishes unless the patient requests it.	There is some evidence of providers taking the time to discuss tradeoffs with patients, however this is inconsistent and not supported by any tools. Fewer than 10% of patients have had advanced care planning discussions which are documented in the patient's medical record.	Physicians talk to patients about the tradeoffs of different approaches of care. Scheduling systems allow for longer appointments to facilitate shared decision making or advanced care planning discussions, or there are other processes in place to facilitate these difficult decisions. This includes decision support tools and advanced care planning tools. Advanced care plans are documented in >20% of patients' medical records.
5. Complaints are gold	Patient feedback is not collected.	Patient feedback is collected and tracked, and used to celebrate successes. Negative feedback is viewed inconsistently, with some recognizing its validity to drive improvement and others dismissing it as idiosyncratic or intractable.	Patient feedback is collected both through structured surveys and more informally (i.e., through mystery shopping). Negative feedback is seen as helpful information for improvement. Exposing and resolving problems is a part of a regular process rather than being the result of extraordinary efforts.

Ordinally scored	Score 1	Score 3	Score 5
6. Comprehensive primary care	Practice offers fewer than 3 of the following services: Intravenous fluids, diuretics and/or antibiotics; Nebulizer treatments; Anticoagulation services; Insulin initiation and/or stabilization; Insulin pumps; Colonoscopies or sigmoidoscopies; Colposcopies/LEEP; IUD insertion; Casting; Suturing; Incision and drainage; Skin biopsies; Mole and skin tag removal; Joint injections.	Practice offers 5-7 of the following services: Intravenous fluids, diuretics and/or antibiotics; Nebulizer treatments; Anticoagulation services; Insulin initiation and/or stabilization; Insulin pumps; Colonoscopies or sigmoidoscopies; Colposcopies/LEEP; IUD insertion; Casting; Suturing; Incision and drainage; Skin biopsies; Mole and skin tag removal; Joint injections.	Practices offers 9 or more of the following services: Intravenous fluids, diuretics and/or antibiotics; Nebulizer treatments; Anticoagulation services; Insulin initiation and/or stabilization; Insulin pumps; Colonoscopies or sigmoidoscopies; Colposcopies/LEEP; IUD insertion; Casting; Suturing; Incision and drainage; Skin biopsies; Mole and skin tag removal; Joint injections.
7. Careful selection of specialists	Referrals are made to a specialty, rather than to a specific clinician. PCPs rarely get timely feedback reports from specialists or hospitals that treat their patients.	There is an implicit preferred network of specialists, who share a practice style and share feedback with the PCP in a timely manner, however this is provider dependent and is based on individual relationships. Costs are considered when making decisions about facilities or specialists, for example, if affordability is an issue for a patient.	The care team uses data to create a "preferred specialist" list on the basis of cost and quality performance. Monthly reports on quality and utilization outside the practice are used to inform decision-making about treatment choices, including which facilities and providers to send patients to.

Ordinally scored	Score 1	Score 3	Score 5
8. Coordination of care	There is no notification when a patient is admitted to the ED or hospital. PCPs rarely get timely feedback reports from specialists or hospitals that treat their patients. There are no systems in place to know if a patient has attended their specialist appointment.	The PCPs sometimes hear about an admission or ED visit from the hospital, but more often than not from family member. There is no automated process for making sure someone attends his or her appointment, that a timely follow-up is in place or that their patient record has been updated accordingly. There is no outreach to bring patients in proactively and following up on care happens inconsistently.	The clinic maintains close relationships with hospitalists and hospital-based care managers to ease transitions. ED physicians may call them before making the decision to admit a patient. There is a system in place to ensure a patient attends his or her appointment and there is a timely transfer of information back to the clinic post specialist visit. The clinic has a process in place for ensuring the patient has a timely follow-up appointment following an ED visit or admission.
9. Upshifted staff roles	Staff members work separately, with little sense of team. Staff often end up undertaking tasks for which they are not qualified or over-qualified when they could be used elsewhere.	Care teams exist, but physicians still spend a lot of time on tasks that do not use their clinical expertise. Protocols are in place for non-clinical staff and are being followed at least 50% of the time.	Physicians are supported by NPs/PAs, RNs, MAs, and/or administrative team members who are working at the "top of their license". Provider skill sets are matched to patient needs.
10. Standing orders and protocols	There is little standardization and no protocols (e.g. different clinical and non-clinical staff have different approaches to the same processes).	Some protocols may have been created, but they may not be commonly used becausre they are not easily available or their use is not adequately monitored. Protocols are in place for non-clinical staff and are being followed at least 50% of the time.	Protocols are known and used by clinical and non-clinical staff and regularly followed-up on through some form of monitoring and oversight. Adherence is at least 85% for established protocols. If someone diverges from a protocol, they have to articulate and record the rationale.

Dichotomously scored	No	Yes	
11. Shared work spaces	Staff do not work in an open-plan environment/bull-pen	Staff work in an open-plan environment/bull-pen	
12. Balanced compensation	Physician compensation is purely based on cash collections or productivity (i.e., RVUs).	Physician compensation is based on at least one of the following: -Measures of performance on quality of care -Patient experience -Resource utilization -Contribution to practice-wide improvement activities	
13. Low overhead on space and equipment	Owned on-site lab or advanced imaging (i.e. CT, MRI).	No capital-intensive ancillary services (e.g. only phlebotomy, flat x-ray, point of care lab testing)	