Methodological Appendix

The quantitative analysis is focused on measuring the impact of the Accountable Care Collaborative on health care utilization, costs and quality. The analysis was performed using data before and after the Accountable Care Collaborative was implemented. The full set of parameter estimates associated with the results reported in the final report are available from the authors upon request.

The results are based on a sample that was stratified based on the time an individual enrolled in the ACC. We divided Accountable Care Collaborative enrollees into three cohorts, defined based on when the enrollees entered the Accountable Care Collaborative. Cohorts 1, 2 and 3 consist of enrollees that joined the Accountable Care Collaborative in FY2011-12, FY2012-13, and FY2013-14, respectively. The control group for each cohort consists of fee-for-service beneficiaries who did not join the Accountable Care Collaborative. We stratified our analysis between enrollees who only qualify for Medicaid coverage using traditional criteria and enrollees who are eligible for Medicaid and Medicare coverage.

A. Controlling for Non-Random Enrollment into the Accountable Care Collaborative

We incorporated the enrollment process into our analysis because Accountable Care Collaborative enrollment is not random. This aspect of our analysis differs from previous analyses of the impact of the Accountable Care Collaborative. We controlled for non-random selection in order to minimize the influence of selection on our results. Enrollment into the Accountable Care Collaborative was initially based on attribution of Accountable Care Collaborative-eligible individuals to primary care providers. During the first year a client who was attributed to a primary care provider that joined the Accountable Care Collaborative as a "PCMP" was enrolled by default. In subsequent years, clients who were attributed to a primary care provider who was ***not*** in the Accountable Care Collaborative were ***not*** automatically enrolled.

Primary care providers that choose to join the Accountable Care Collaborative may do so for reasons that are potentially correlated with spending and outcomes. For example, a provider who is already a believer in and perhaps a practitioner of "Accountable Care" may be more apt to join the Accountable Care Collaborative. This same provider may have already been performing well on the KPIs and may have care management processes. For similar reasons, providers who do not join may be different that providers that do join. Failure to control for this would lead us to over-estimate savings related to the Accountable Care Collaborative because of selecting providers, rather than actually changing how care is delivered. By controlling for selection of providers we can estimate the extent that the Accountable Care Collaborative led to changes in the way care is delivered rather than differences among providers in practice style.

We incorporated this enrollment mechanism into our analysis by annually attributing all Accountable Care Collaborative-eligible clients to primary care providers during the entire sample period. Information on the attributed primary care providers (clinic type and specialty) and client demographics (age, race, primary language, and Medicaid enrollment length of time) was used to calculate the probability, or propensity, that an individual would join the Accountable Care Collaborative. The predicted propensities were then used as weights to make the FFS control group more similar to the Accountable Care Collaborative enrolled group based on observable characteristics. Persons in FFS who had more similar characteristics to the Accountable Care Collaborative enrollees were weighted more heavily than individuals who had characteristics that were dissimilar to the Accountable Care Collaborative enrollees. The propensity score weights specifically control for client and provider characteristics that influence a client's choice of provider (Guo and Fraser, 2014). We also estimated a specification with attributed provider fixed effects to control for the provider's decision to join the Accountable Care Collaborative as a sensitivity analysis. Because the inclusion of fixed effects did not affect our estimates we reported the results of the specification that did not include provider fixed effects in the final report.

A critical step in this analysis is to replicate the attribution algorithm used by SDAC and apply it to everybody in the dataset. Our "pseudo-attribution" and the actual attribution is described in Table 1. The approaches are slightly different because of available data but the differences are unlikely to have a meaningful impact on our results because our attribution uses exogenous characteristics to break ties. First, we replicated the attribution algorithm used to attribute individuals to a primary care medical provider (PCMP). Individuals who were eligible for the Accountable Care Collaborative were attributed to a primary care medical provider based on prior twelve months of evaluation and management visits at primary care providers. If clients had a majority of visits at a provider then the client was attributed to the provider. In this step our "pseudo-attribution" matched the actual attribution 75% of the time, including ties[1]. In the actual attribution ties were broken using other types of utilization, attribution of other family members, or the most recent visit. We were unable to perform attribution based on family members because we did not have information about family relations in our dataset. Individuals without prior E & M utilization were indicated as unattributed.

B. Methodology

We performed a propensity-score weighted difference-in-differences analysis on each cohort and eligibility type (Ryan, Burgess and Dimick, 2014; Stuart, *et. al*, 2014, Dimick and Ryan, 2014) using Stata Version 14. The sample consisted of individuals who were continuously enrolled and eligible to be in the Accountable Care Collaborative during a given quarter (i.e. three month period). We excluded utilization during an adjustment period of one quarter before and after the quarter of enrollment into the Accountable Care Collaborative because preliminary analyses revealed a spike in utilization during the three months after the first three months of enrollment.

[1] Matches are defined as either one-to-one matches or, in the case of ties, we identified the actual PCMP among the set of primary care providers that were tied. We did not consider individuals who were not attributed to PCMPs in our comparison.

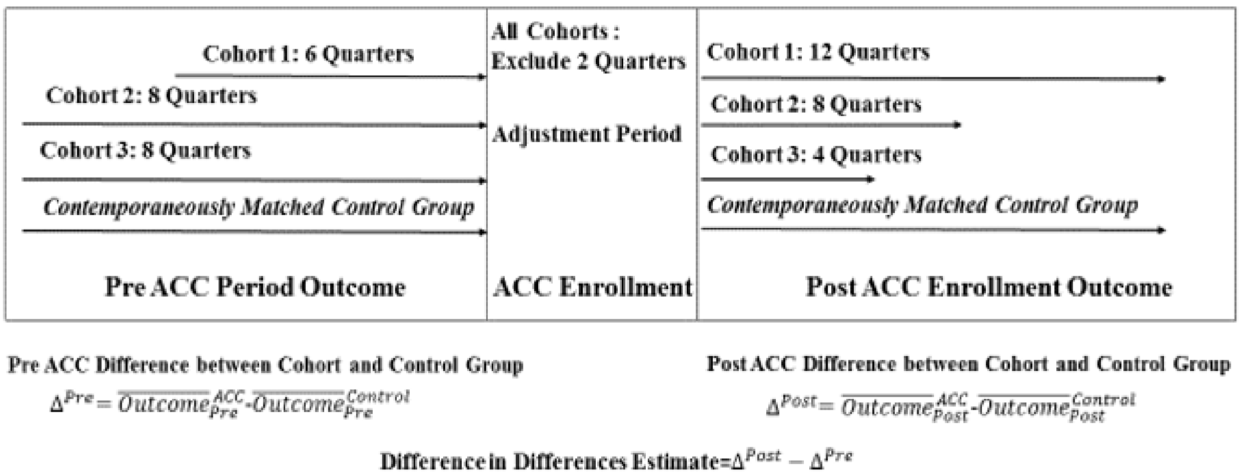
Table 1. Comparison of Actual attribution to the "Pseudo-attribution"			
		Pseudo-Attribution	Actual Attribution
Step 1	Prior 12-month Evaluation and Management Utilization at Primary Care Medical Providers as described in the 3M/Treo Solutions "COSDAC Client Attribution and Enrollment Processing" document	Identical Approach	Identical Approach
Step 2	If Single Provider identified →	Attribute	Attribute
Step 3	If multiple providers are tied, attribute using→	Most recent utilization	Use attribution of family members, most recent utilization, or other utilization to attribute
Step 4	If no attributable providers were identified in previous 12 months →	Unattributed	Unattributed

The study design is summarized in Figure 1. The length of the pre-period for cohort 1 is restricted by available data and is at most 6 quarters long. The pre-period is longer for Cohorts 2 and 3 because they started in FY2012-13 and FY2013-14, respectively. The length of the post-period follow-up is up to four years (16 quarters) for Cohort 1 and is shortened by one year for each successive cohort. We calculate the difference-in-differences estimate by subtracting the Accountable Care Collaborative and Control group difference in post-period adjusted outcomes from the Accountable Care Collaborative and Control group difference in pre-period adjusted outcomes as indicated at the bottom of Figure 1.

Propensity scores were estimated using Accountable Care Collaborative enrollees in the year prior to enrollment as the reference group. The predicted propensity scores were used to compute inverse probability weights based on an Average Treatment Effect of the Treated (ATET) specification. Separate weights were computed for the Accountable Care Collaborative group one year pre-Accountable Care Collaborative versus: two years pre-Accountable Care Collaborative, one, two, three and four years post-Accountable Care Collaborative, respectively. The control group weights also used the one pre-Accountable Care Collaborative population as the reference category for the control group weights one-year pre-Accountable Care Collaborative control group and the equivalent control group versions of the respective Accountable Care Collaborative weights.

We estimated the difference-in-difference analysis using total spending and spending on inpatient, outpatient, and pharmaceuticals. We also analyzed the following subcategories of spending: evaluation and management; outpatient emergency department; outpatient hospital; radiology; and laboratory. We also analyzed quality and utilization used measures derived from

Figure 1. Study Design



the administrative claims data. These are listed in Table 2. A subset of these measures are closely related to the Key Performance Indicators (KPIs) but differ in the inclusion criteria used to compute the sample "exposed" to the utilization/outcome. We also calculated the KPIs but also estimated the versions in Table 2 for comparability to other studies. The performance on KPIs has been under close scrutiny because they are tied to bonus payments and the trends over the sample time period are well-documented in HCPF Annual Reports. The measures chosen include utilization on low-value or unnecessary care as well as utilization of high-value preventative care that may be related to Accountable Care Collaborative or compliance of the patient.

The prevalence of zero spending led us to estimate a "two-part" model where the probability of any use was estimated using a logistic regression in the first part. A generalized linear model with a log-link function was used in the second part. The log-link function was necessary because health care expenditure data is highly skewed. We controlled for a comprehensive set of risk-adjustors calculated using diagnoses made in the previous 3 months; age categories (infants, toddlers aged 1-4, children aged 5-14, teens and young adults aged 15-24, adults aged 25-44, and adults aged 45 – 64). We also included gender (Male=1, Female-0) and gender interacted with the latter three age categories using the Chronic Illness and Disability Payment System (CDPS) Version 6.0 grouper. The CDPS was developed to use in adjusting capitated payments for Medicaid beneficiaries by Dr. Richard Kronick, Dr. Todd Gilmer and colleagues at the University of California, San Diego (http://cdps.ucsd.edu/index.html). This grouper is freely available and uses many of the same elements as the proprietary system sold by 3M. We used diagnoses from the previous quarter to adjust spending and utilization for the current quarter. We also controlled for the clients' race and language spoken because these may be correlated with both utilization and Accountable Care Collaborative enrollment. Attributed provider type (FQHC, RHC, Pediatric, OBGYN, hospital-based clinic; reference group: freestanding private primary care clinic) and the strength of the attribution measured as the share of previous 12 month E&M visits at the attributed provider were also included in the specification. We estimated the spending model with and without attributed provider fixed effects to examine whether the propensity score approach sufficiently controlled for selection of providers in the Accountable Care Collaborative. We settled on the specification without fixed effects because it is are more efficient and because inclusion of fixed effects had no bearing on neither our specification tests nor the estimates.

We performed tests to assess the validity of the propensity score specification including overlap and balancing. In addition, we tested the difference between the Accountable Care Collaborative and Control group pre-period trends. The accuracy and interpretation of difference-in-difference estimates is reliant on parallel trends of the Accountable Care Collaborative and control groups in the pre-period. The Cohort 1 specification of the adult and children samples passed the test for parallel trends. However, the Cohort 2 and Cohort 3 samples failed the test of pre-period parallel trends. As a result, we included a treatment group and trend interaction in the final specification to allow for separate Accountable Care Collaborative and FFS trends over the sample period.

The specification of the quality and utilization measures were estimated using a logistic regression using the same propensity score weights that were used in the spending specifications. The outcome measures and the source are listed in Table 2.

C. Data

We analyzed paid administrative claims submitted for services rendered to Medicaid beneficiaries between July 2009 and June 2015. We defined total spending and inpatient, outpatient, pharmaceutical and other spending using place of service, claim type, CPT codes, and category of service codes. Within each respective subcategory we looked at several types of services defined using procedure codes (CPT).

To enter the sample, clients must have been continuously enrolled in Medicaid with either AFDC, Foster Home, and BC Women and Children to be classified as a "Standard" enrollee, or dually eligible for Medicare Part A, B, or both to be classified as a MME enrollee. In addition, subsequent utilization is included if individuals have been continuously enrolled in either FFS or the Accountable Care Collaborative during a given quarter to be included in the sample. This restriction leads us to exclude utilization if the individual enrolled in the middle of the quarter. We do so to ensure an Apples-to-Apples comparison between the treatment and control groups.

The spending was first categorized by inpatient, outpatient, and pharmaceutical claims. Next, we separately calculated outpatient spending in: community or clinic; hospital outpatient; and hospital ED setting based on the place of service. We further categorized spending into Evaluation and

Management (E&M)[2]; Medicine[3]; Surgery (CPT: 00001-69999); Radiology (CPT: 70000-79999); Pathology and Laboratory tests (CPT: 80000-89999); and Other which includes all HCPCS codes and all other spending, including transportation and medical supplies.

Table 2. Quality and Utilization Measures

Measure Name	Definition	Source
Potentially Avoidable ED Visits, 18+	ED visits with a primary diagnosis indicating they are "potentially avoidable"	1
Potentially Avoidable ED Visits, 1-17	ED visits with a primary diagnosis indicating they are "potentially avoidable"	1
Ambulatory ED utilization	Ambulatory ED visits (that don't result in admission)	1,2
Avoidance of CT without Ultrasound, for Evaluation of Suspected Appendicitis	Had a CT scan, but NOT an ultrasound, within 30 days prior to index case	1
Use of Appropriate Medications for People With Asthma, Adult	Members who were appropriately prescribed (and filled) medication during the measurement year	4,5
Use of Appropriate Medications for People With Asthma, Children	Members who were appropriately prescribed (and filled) medication during the measurement year	4,5
Adolescent Well-Care Visits	Had at least one comprehensive well-care visit during the measurement year	2,3
Diabetes HbA1C Testing	Had at least one HbA1c test performed during the measurement year	2,3, 5
Appropriate Testing for Children with Pharyngitis	Received a group A streptococcus (strep) test for the episode	2,3,4,5
Avoidance of Head Imaging for Uncomplicated Headache	Members that had a CT or MRI for an uncomplicated headache	1
Appropriate Use of Imaging Studies for Low Back Pain	Did not receive an imaging study within 28 days of diagnosis	1,4,5
Annual Monitoring for Patients on Persistent Medications	Had at least one therapeutic monitoring event for the therapeutic agent in the measurement year	5
Adult Prevention Quality Overall Composite	PQI Overall Composite; readmits counted twice	1, 6
Adult Prevention Quality Acute Composite	PQI Acute Composite; readmits counted twice	1, 6
Adult Prevention Quality Chronic Composite	PQI Chronic Composite; readmits counted twice	1, 6
Well-Child Visits for Children 0-15 Months of Age	Had 6 or more well-child visits during their first 15 months of life	1
Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life	Had one or more well-child visits during the measurement year	1

Notes: 1. Related to KPI; 2. Oregon CCO Accountability Measure; 3. CMS; 4. PQRS; 5. NIH; 6. Prevention Quality of Care indicator based on preventable inpatient utilization.

[2] E&M CPT codes: 92002, 92004, 92012, 92014; 99201-99215; 99241-99245; 99304-99316; 99318; 99324-99328; 99334-99337; 99341-99350; 99381-99387; 99391-99397; 99401-99404; 99411-99412; 99420; 99429; and 99461.
[3] Medicine CPT codes: 9#### but not an E&M code

Comparisons of the standardized differences between the FFS and Accountable Care Collaborative groups were made to assess the balance of covariates, with and without propensity score weighting. The specification passed standard thresholds of <0.10 difference between standardized means and variance ratios between 0.90 – 1.10. Note that the samples are unbalanced because members can cycle on and off Medicaid during the sample period. Overall, the number of clients in our sample is smaller than the entire population because of the inclusion of an adjustment period and the requirement of at least six-months of continuous enrollment.

References:

Dimick, J. B., & Ryan, A. M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*, *312*(22), 2401-2402.

Guo, SY. And Fraser, MW (2014) *Propensity Score Analysis: Statistical Methods and Applications, Edition 2.* Vol. 11. Sage Publications, 2014. ISBN 13: 978-1-45223-500-4

Ryan, Andrew M., James F. Burgess, and Justin B. Dimick. "Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences" *Health Services Research* (2014).

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services and Outcomes Research Methodology*, *14*(4), 166-182.