Effective Care Management by Next Generation Accountable Care Organizations

Neal O'Hara, BA; Olivia C. Tran, MPH; Shantanu Phatakwala, MS; Anita Cattrell, PhD; and Youssef Ajami, MS

As of 2019, the 518 accountable care organizations (ACOs) participating in the Medicare Shared Savings Program (MSSP) coordinate care for 10.9 million fee-for-service (FFS) beneficiaries.[1] More than 70% currently participate in upside-only tracks, but the Pathways to Success final rule requires that all future MSSP ACOs take on some level of downside risk within 2 years.[2] Financial performance for participants in the 2-sided Next Generation ACO (NGACO) model varied widely, ranging from owed losses of $14 million to earned savings of nearly $30 million in 2017,[3] highlighting the need for reproducible strategies that can improve outcomes and reduce spending.

Most NGACOs implemented care management programs targeted at subpopulations with chronic comorbid diseases as a way to manage costs,[4] but there are relatively few examples in the literature that demonstrate an association between complex care management programs and lower medical expenditures for participating beneficiaries.[5-11] Findings from recently published success stories were limited to a single ACO site or based on a small sample size of Medicare FFS patients, which limits generalizability.[7,11] To our knowledge, there are no large-scale, multisite ACO studies that evaluate the impact of a standardized complex care management program on cost and utilization outcomes.

Some previous studies have identified care management strategies that may be associated with improved outcomes, such as proactive targeting of populations with preventable risk factors,[7,9,12-14] evidence-based intervention design,[9,11,15] and continuous performance management facilitated by key operational indicators.[16] However, there is little detail regarding the relative importance of implementing specific program features[17] or the impact of intervention fidelity on program effectiveness. Fidelity, or the degree to which an intervention is applied as intended, is generally acknowledged as a modifier of population health program effects in literature,[5] but its precise impact on program effectiveness is not well understood.

AJMC.com
Copyright © Managed Care & Healthcare Communications, LLC

ABSTRACT

OBJECTIVES: The objectives of this study were to estimate the utilization and spending impact of a standardized complex care management program implemented at 5 Next Generation accountable care organizations (NGACOs) and to identify reproducible program features that influenced program effectiveness.

STUDY DESIGN: In 2016 and 2017, high-risk Medicare beneficiaries aligned to 5 geographically diverse NGACOs were identified using predictive analytics for enrollment in a standardized complex care management program. We estimated the program's impact on all-cause inpatient admissions, emergency department visits, and total medical expenditures (TME) relative to a matched cohort of nonparticipants. In a subanalysis, we studied the modifying effects of intervention fidelity on program impact.

METHODS: We created 1897 propensity score–matched case-control pairs based on preprogram similarities in disease profile, predictive risk score, medical cost, and utilization. Changes in outcomes 6 months post program were measured using difference-in-differences analyses. We used principal components analysis to identify program features associated with reduced inpatient admissions, classified cases according to intervention fidelity, and measured postprogram changes in TME for each subgroup.

RESULTS: Program participation was associated with a 21% reduction in all-cause inpatient admissions ($P = .03$) and a 22% reduction in TME ($P = .02$) 6 months after program completion. Relative spending reductions were 2.1 times greater for high-fidelity interventions compared with overall program participation ($P < .001$).

CONCLUSIONS: Centrally staffed complex care management programs can reduce costs and improve outcomes for high-risk Medicare beneficiaries. Integrating predictive risk stratification, evidence-based intervention design, and performance monitoring can ensure consistent outcomes.

Am J Manag Care. 2020;26(7):296-302. doi:10.37765/ajmc.2020.43759

OBJECTIVES

The principal objective of this study was to estimate the clinical and financial impact of a standardized complex care management program implemented across 5 geographically diverse NGACOs during 2016 and 2017. We analyzed whether program participation was associated with changes in all-cause inpatient admissions, emergency department (ED) visits, and total medical expenditures (TME) for participating beneficiaries. In a subanalysis, we identified specific program features that were associated with improved outcomes, developed a generalizable measure of intervention fidelity, and estimated the modifying effects of intervention fidelity on program outcomes.

TAKEAWAY POINTS

This large-scale, multisite study demonstrates that a standardized complex care management program can consistently reduce utilization and spending for high-risk Medicare populations distributed across diverse geographies.

- Program participation was significantly associated with reduced all-cause inpatient admissions (–21.2%) and lower total medical expenditures (–22.0%) compared with a propensity score–matched cohort of nonparticipants.
- We identified specific program features that were significantly associated with intervention fidelity.
- Relative spending reductions were 1.9 times greater for high-fidelity interventions compared with overall program participation.
- Future accountable care organization leaders can use these findings to inform effective care management program design.

METHODS

Overview of Complex Care Program

In 2016 and 2017, Evolent Health, a provider of value-based care services, partnered with 5 NGACOs to implement a complex care management program targeted at Medicare beneficiaries with chronic comorbidities and a high risk of hospitalization.[13] In the program, registered nurses led an evidence-driven, team-based care advising approach[9,18-22] aimed at reducing hospitalizations, ED visits, and TME among participating beneficiaries. Common chronic conditions within the target population included chronic obstructive pulmonary disease, coronary artery disease, diabetes, congestive heart failure, and asthma. Some patients also exhibited comorbid diagnoses with chronic kidney disease, behavioral health conditions, or neurocognitive disorders.

Enrollment

Nearly 87% of program participants were proactively identified via predictive risk stratification. Machine learning models predicted each patient's risk of incurring a future avoidable hospitalization using administrative, clinical, and sociodemographic variables such as comorbid diagnoses, condition severity, acute utilization trends, laboratory values, educational attainment, and food access.[13,23,24] A small subset of enrollees (13%) were referred to the program by their physician.

Once identified, patients were added to a queue for telephonic outreach from a registered nurse. Program coordinators prioritized patients for outreach according to predicted admission risk, along with the output from a separate predictive model that estimated each prospective participant's likelihood of enrolling in the program.[25] Across all ACOs during the study period, approximately 3% of patients were identified as appropriate for the program.

Program Design

Registered nurses collaborated with patients, their physicians, and an extended care team to develop an individualized care plan focused on 6 key mechanisms: barrier identification and action planning, gap in care closure, care coordination, basic medication reconciliation,[22] patient activation and education,[19-21] and referrals to local or electronic resources. Program duration averaged 4 months, and patient-care advising interactions were designed to occur at least every 14 calendar days. More than 96% of patient-care advising interactions occurred telephonically. Patients graduated from the program once all identified barriers were resolved.

A web-based care management workflow tool was used to document all program activities, monitor key performance indicators, trigger relevant clinical alerts, and identify patients due for follow-up. Program coordinators periodically reviewed aggregate performance data each month to identify best practices and develop process improvement plans.

Study Population

Intervention and control samples were created from a pooled population of 163,977 Medicare beneficiaries aligned to 5 NGACOs in California, Idaho, Illinois, Indiana, and Virginia. The intervention sample includes beneficiaries 65 years and older who enrolled in complex care management between January 1, 2016, and December 31, 2017. All patients who participated in the program and had no other care management program enrollment in the 12 months preceding identification were eligible for inclusion in the intervention sample. Patients who were identified as appropriate for care management via predictive risk stratification but declined to participate, could not be reached for initial enrollment, or were not contacted during the study period because of program capacity constraints served as candidate controls. In separate sensitivity analyses, we restricted the control group to include only patients who were identified via predictive stratification but not contacted during the study period.

Data Sources

Program operations data were used to identify intervention and control patients. Outcome measures were derived from administrative claims data and reported in terms of rates per member per month or per 1000, where applicable. The baseline period refers to the 12-month period prior to the program start date for intervention

TABLE 1. Baseline Characteristics and Utilization Among Control and Intervention Groups Before and After Propensity Score Matching

	Baseline			After matching		
	Control	Intervention	*P*	Control	Intervention	*P*
Patients, n	3666	1993		1897	1897	
Age in years, mean	77.7	76.8	<.001[b]	76.9	76.9	.933
Female sex	57%	60%	.021[a]	59%	59%	>.999
Charlson Comorbidity Index score, mean	5.59	5.55	.722	5.55	5.52	.777
Comorbid diseases						
Diabetes	54%	57%	.090	58%	55%	.140
Asthma	17%	20%	<.001[b]	20%	20%	.902
Coronary artery disease	56%	54%	.049[a]	55%	53%	.345
Chronic obstructive pulmonary disorder	39%	39%	.983	39%	39%	.842
Congestive heart failure	48%	46%	.081	46%	45%	.415
Dementia	17%	10%	<.001[b]	9%	10%	.318
Risk score percentile	79.7	79.4	.567	78.3	79.0	.278
Utilization metrics (per 1000 members per month)						
Total inpatient admissions	2548.6	1845.5	<.001[b]	1979.3	1854.2	.162
Acute admissions	1888.2	1445.3	<.001[b]	1542.9	1453.1	.200
Nonacute admissions	639.1	389.3	<.001[b]	417.4	393.3	.470
ACSC admissions	801.2	670.6	<.001[b]	714.8	668.9	.209
ED visits	1462.5	1314.8	.075	1290.7	1294.2	.962
PCP visits	9387.5	9827.1	.091	9458.3	9633.9	.524
Specialist visits	10,187.1	10,300.2	.749	10,316.1	10,224.8	.796
TME (in $ PMPM)	3868.6	2952.0	<.001[b]	3167.7	2963.9	.143

ACSC, ambulatory care–sensitive condition; ED, emergency department; PCP, primary care physician; PMPM, per member per month; TME, total medical expenditures.

[a]$P < .05$.

[b]$P < .001$.

patients and the date of initial identification for controls. Members were followed for 6 months to evaluate utilization and cost outcomes. All analyses were performed in R 3.5.1.

Propensity Score Matching

Across several preintervention periods, we observed that trends were not parallel for the primary outcome of interest, TME, in the treatment and control groups (**eAppendix A** [eAppendices available at **ajmc.com**]), which could introduce bias due to regression to the mean (RTM).[26,27] To mitigate potential RTM bias, we used 1:1 propensity score matching[28] using covariates related to both outcome and assignment[27,29] to select for controls who most resembled patients in the intervention group. We constructed separate multivariable logistic regression models for each ACO population using patient demographics; presence of congestive heart failure, chronic obstructive pulmonary disease, diabetes mellitus, dementia, and other comorbidities; Charlson Comorbidity Index score[30]; predictive risk score; baseline spending; and baseline utilization. The variables used in the propensity score model are listed in **Table 1**. Patients within each ACO were matched using a caliper width of 0.2 times the SD of the propensity score without replacement. Matched pairs were then pooled into a final study data set. We assessed covariate balance both within each ACO and across the entire study cohort.

Difference-in-Differences Analysis

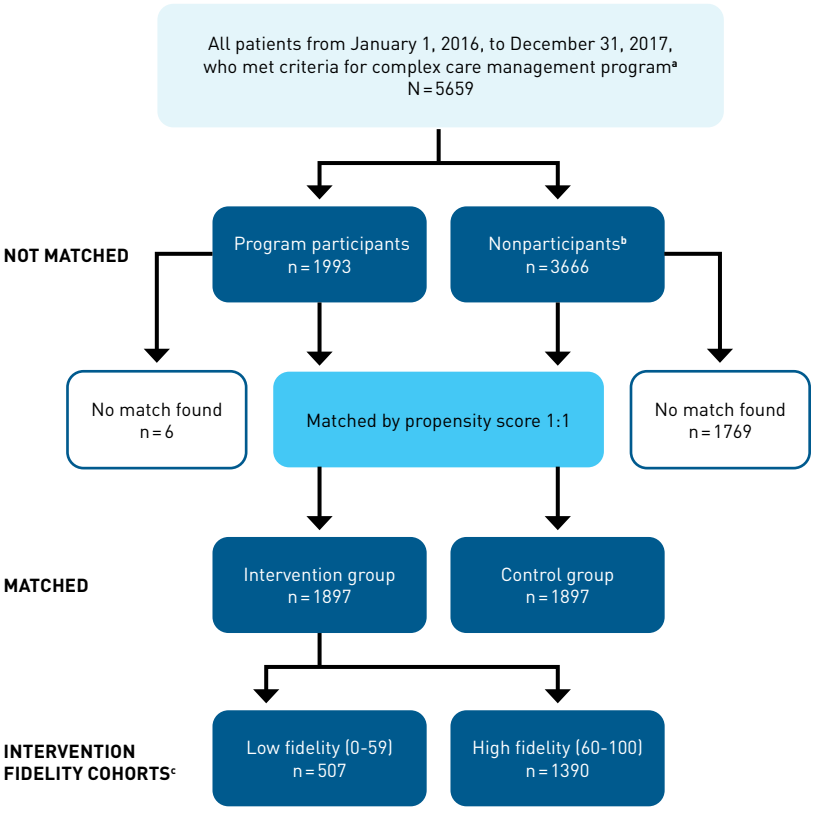
Following propensity score matching, our primary analysis utilized a difference-in-differences (DID) approach to compare the change in cost and utilization outcomes before and after the program among the intervention group with the change in the control group over the same period.[31] The outcomes of interest were all-cause inpatient admissions, ED visits, and TME. We used linear probability models that adjusted for month of year, years of ACO experience, and ACO site.

Intervention Fidelity Analysis

Using program operations data collected by the workflow management tool, we explored modifiable program components related to patient identification and care management intervention delivery that were associated with reduced all-cause inpatient admissions post program. Using principal components analysis (PCA) and multivariable regression techniques, we isolated 5 reproducible program features that were significantly associated with reduced admissions.[32,33] We created binary flags for each feature and assigned each case in the intervention sample a normalized score of

intervention fidelity from 0 to 100. The intervention fidelity score was calculated by observing whether each feature was present during the study period and weighted according to the estimated percentage of variance explained by each feature. Using sensitivity analyses, we determined intervention fidelity score thresholds and classified each intervention participant into low- and high-fidelity subgroups. The score was validated using a separate cohort of program participants to ensure replicability and generalizability.

FIGURE 1. Study Population Criteria



[a]Medicare beneficiaries aligned to the 5 Next Generation accountable care organizations in the study cohort with a high risk of hospitalization and at least 9 months of data in the preintervention period and 3 months of data in the postintervention period.

[b]Reasons for nonparticipation include patient declined to participate (n = 2208), patient contacted but not reached (n = 778), and patient not contacted during the study period because of program capacity constraints (n = 680).

[c]Intervention fidelity refers to the degree to which an intervention is implemented as intended. A score of 0 to 100 was assigned to each case in the intervention sample using a weighted equation based on the observed occurrence of key program activities during the study period (Figure 2).

RESULTS

During the study period, 1993 total patients enrolled in the complex care program and had sufficient historical and follow-up data available (**Figure 1**). Six patients were excluded from the analysis because we did not identify an appropriate matched control. The final analysis included 1897 1:1 matched pairs. Prior to matching, we found significant differences between the intervention and control groups that were balanced following propensity score matching (Table 1). After matching, we observed approximately parallel preperiod trends in both the intervention and control cohorts for the primary outcome, TME (**eAppendix B**), satisfying the parallel trend assumption required for internal validity of DID estimates.

Utilization and Spending

Using regression analysis, we estimated the adjusted predicted change in all-cause inpatient admissions, ED visits, and TME for the intervention group relative to the change observed during the same period in the control group (**Table 2** [A]). The models adjusted for potential year-over-year changes in cost and utilization patterns. Compared with the control group, participation in the complex care program was significantly associated with a 21.2% decline in inpatient admissions (95% CI, −37.1% to −5.4%; $P = .03$) and a 22.0% reduction in TME (95% CI, −37.6% to −6.5%; $P = .02$). We also observed a relative 3.0% decline in ED visits for program participants (95% CI, −20.1% to 14.2%), but the difference was not statistically significant ($P = .78$).

We also conducted sensitivity analyses that restricted the control group definition to include only patients who were in queue for outreach, and we observed that inpatient admissions were reduced by 21% and TME was reduced by 17% (**eAppendix C**). The magnitudes of the reductions were similar to the results in the unrestricted control group, but the results are not statistically significant.

Intervention Fidelity

We identified 5 program components during PCA that were associated with intervention fidelity: patient identification based on predictive stratification; graduation from the program with goals achieved; at least 2 patient-care advising interactions per month; at least 1 documented care plan within 14 days of program enrollment; and in-person visits from care managers (**Figure 2**). Each program participant was assigned a normalized intervention fidelity score between 0 and 100 based on observed occurrences of each of the 5 program components identified during PCA and weighted by their estimated relative influence on postprogram reductions in inpatient admissions. After sensitivity analyses, we created 2 subgroups based on the calculated measure of intervention fidelity: low fidelity (0-59) and high fidelity (60-100).

In separate DID models, we observed magnified program effects among program participants in the high intervention fidelity

TABLE 2. Relative Change in All-Cause Inpatient Admissions, ED Visits, and TME

A. Overall care management program participation						
Control (n = 1897) Intervention (n = 1897)	Baseline[a]		Follow-up[b]		Difference-in-differences[c]	
	Control	Intervention	Control	Intervention	Percentage points (95% CI)	*P*
Total inpatient admissions	1979.3	1854.2	2295.3	1743.6	−21.2 (−37.1 to −5.4)	.03
ED visits	1290.7	1294.2	1319.6	1288.2	−3.0 (−20.1 to 14.2)	.78
TME	3167.7	2963.9	3691.2	2866.1	−22.0 (−37.6 to −6.5)	.02
B. Subset of program participants who received high-fidelity intervention (score ≥ 60)[d]						
Control (n = 1390) Intervention (n = 1390)	Baseline[a]		Follow-up[b]		Difference-in-differences[c]	
	Control	Intervention	Control	Intervention	Percentage points (95% CI)	*P*
Total inpatient admissions	1363.5	1298.6	1759.8	841.2	−46.9 (−65.3 to −28.5)	<.001
ED visits	987.8	1006.1	1002.7	938.0	−8.1 (−32.0 to 14.8)	.56
TME	2070.3	2030.4	2747.7	1612.9	−42.2 (−59.5 to −24.8)	<.001

ACO, accountable care organization; ED, emergency department; TME, total medical expenditures.

[a]Defined as 12 months prior to program enrollment or in-queue status.

[b]Defined as 6 months following program enrollment or in-queue status.

[c]Represents the relative change in each outcome for intervention group patients; estimates are from linear probability models adjusting for month of year and ACO program year as fixed effects and ACO as an indicator variable.

[d]Intervention fidelity refers to the degree to which an intervention is implemented as intended. A score of 0 to 100 was assigned to each case in the intervention sample using a weighted equation based on the observed occurrence of key program activities during the study period (Figure 2).

FIGURE 2. Program Features Associated With Reduced Inpatient Admissions, Ordered by Relative Influence Determined in Principal Components Analysis

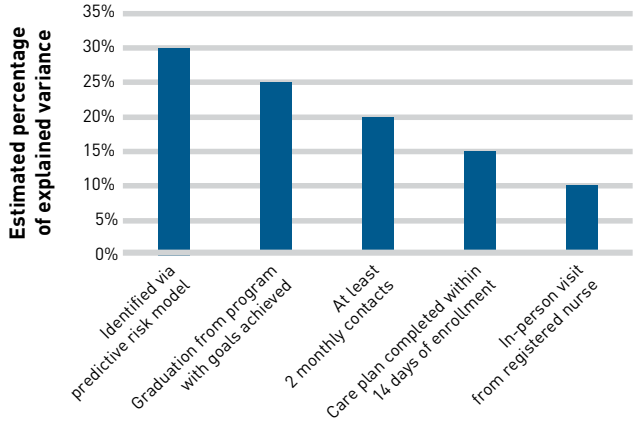
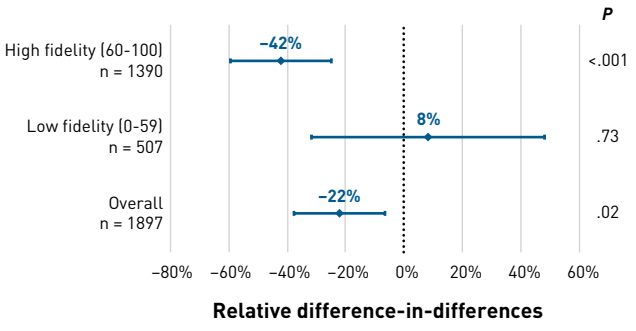


FIGURE 3. Relative Changes in Total Medical Expenditures Associated With Participation in Care Management, Overall and by Intervention Fidelity Score[a]



[a]Intervention fidelity refers to the degree to which an intervention is implemented as intended. A score of 0 to 100 was assigned to each case in the intervention sample using a weighted equation based on the observed occurrence of key program activities during the study period (Figure 2).

subgroup (60-100) compared with the low-fidelity subgroup (0-59) and the mean reduction in TME for all program participants in the intervention sample (**Figure 3**). Within the high-fidelity subgroup, inpatient admissions were reduced by 46.9% (95% CI, −65.3% to −28.5%; $P < .001$) and TME was reduced by 42.2% (95% CI, −68% to −25%; $P < .001$) 6 months post program, which is 1.9 times greater than the mean TME reduction observed among all program participants in the study. We also observed an 8% reduction in TME among patients in the low intervention fidelity subgroup, but the results were not statistically significant (95% CI, −31.6% to 48.2%; $P = .73$) (Table 2 [B]).

DISCUSSION

To our knowledge, this analysis is the first large-scale study to demonstrate that a centralized complex care management program can reduce total medical expenditures among high-risk beneficiaries across multiple ACO sites. Sensitivity analyses showed consistent program impact within each NGACO and across the entire study population. Given that the NGACOs in the study were heterogenous in terms of size (ranging from approximately 14,000 to 45,000 beneficiaries), geography, and experience with the NGACO model (3 of the 5 were in the first year of participation in the program), this study's findings may be generalizable to other MSSP ACOs.

To further evaluate our findings, we identified studies that found an association between high-risk care management programs and reduced utilization and spending for participants. The magnitude of the reduction in hospitalizations and ED visits that we observed among complex care program graduates is within previously reported ranges.[7,9,11] However, compared with other studies of successful high-risk Medicare care management program implementations, our analysis found that program participation reduced medical expenditures to a greater extent. Our findings show that TME for program participants decreased by 22% relative to nonparticipants. Other studies that found lower spending among program participants reported relative reductions of 6% to 9%.[7,11]

We hypothesize that several factors could account for the relatively higher program effects observed in our study. First, we evaluated a mature program that was staffed by teams with previous implementation experience. Therefore, our results may partly reflect the time required to develop effective care management and patient engagement strategies, which was noted as a potential determinant of success in previous studies.[8,9]

Second, we used machine learning models that employed diverse clinical, administrative, and sociodemographic data sources to identify prospective program participants who are likely to incur a future avoidable hospitalization. During our literature review, we found that other successful programs used less sophisticated targeting criteria based on the presence of chronic comorbidities[5] or prior utilization.[7,11] Results of previous studies have shown that predictive models can accurately predict future preventable utilization[13,23,24] and produce superior results relative to condition-based criteria.[34] However, although predictive risk stratification has been widely adopted by ACOs, many organizations rely solely on claims-based models[34,35] and do not leverage other clinical data sources that may predict future risk. Our findings suggest that predictive risk stratification algorithms, particularly models that include diverse data sources, are a critical component of program effectiveness. Moreover, analytics that inform patient engagement strategies can help ensure that programs achieve maximum reach.[25,36]

Finally, our intervention design was grounded in evidence-based best practices[17-25] and was continuously monitored throughout implementation using a centralized care management workflow tool. Leveraging real-time program operations data, program coordinators were able to monitor specific care-advising activities, track participant adherence to recommendations, and identify opportunities to improve implementation. We observed that postprogram spending reductions among cases in the high-fidelity subgroup, in which patients consistently received program features associated with reduced risk of hospitalization, were 1.9 times greater than the reductions observed among overall program participants (42% vs 22%). Of the 1897 program participants in the intervention group, 1390 (73%) received high-fidelity interventions, which suggests that the consistent application of high-value program components was a significant driver of program success.

NGACO leaders reported difficulties implementing performance management strategies, citing infrequent tracking that occurred on a monthly and sometimes quarterly basis.[4] We hypothesize that the weekly monitoring cycle employed by the care management staff in this study contributed to the program's effectiveness. As has been found with the management of chronic disease prevention needs,[37,38] real-time data feeds were critical to the performance management process and provided care managers with timely information that influenced patient-level care.

Limitations

It is important to acknowledge potential limitations of our study. For example, enrollment in the complex care program was nonrandom, which introduced the potential for biased DID estimates. We took several steps to control for measurable confounding and to mitigate potential bias due to regression to the mean during the propensity score–matching process. For example, we matched the cases to controls who were similar across several baseline characteristics,[7-11] including preperiod outcomes and time-invariant covariates such as age and sex (**eAppendix D**).[26-29] Given the numerous variables that were examined for the propensity score models, we assumed that any unmeasured confounders would be distributed similarly across our comparison groups.[39] Moreover, we believe that using the same risk stratification algorithm to identify both treatment and control group subjects mitigated the risk of selecting treatment and control groups that are meaningfully different. To further validate our results, we performed sensitivity analyses that restricted the control group to include only patients who were in queue for outreach but not contacted (eAppendix C). We observed that inpatient admissions and TME were reduced and that the magnitude of the reductions was similar to the results in the unrestricted control group. The results of the sensitivity analysis were not statistically significant, however, and the reduced sample size affected our ability to study the modifying effects of intervention fidelity on program outcomes.

Several areas require further investigation. For instance, more research is needed to understand the specific impact of integrating diverse data sources, including real-time clinical data, on risk stratification accuracy and overall program effectiveness. Similarly, future studies could examine how factors like patient activation,[25,36] provider engagement, and previous organizational experience with care management could affect program success. Finally, this study did not evaluate program implementation costs, calculate return on investment, or quantify overall savings that can be attributed to complex care management. Future studies could evaluate the cost-effectiveness of different complex care management program components.

CONCLUSIONS

For provider organizations considering performance-based risk arrangements, it is important to understand specific tactics that can improve utilization outcomes and reduce medical spending.

This study adds to a growing body of evidence that complex care management programs can reduce all-cause inpatient admissions and TME for participating high-risk Medicare beneficiaries with complex comorbidities.[7,11] We found that program effects were consistent across a relatively heterogenous cohort of NGACOs, suggesting that program effects are reproducible across varying populations.

We identified several program components that are associated with intervention fidelity, such as predictive risk stratification, timely care planning, frequent patient-care advising interactions, and in-person visits. Our findings suggest that program participants who are proactively targeted via predictive stratification and consistently engaged with high-value program activities experience postprogram cost reductions that are nearly 2 times greater than average (Figure 3). Current and future ACO leaders can leverage the findings from this study to implement specific program activities that are associated with program success. ■

Acknowledgments

The researchers and care managers who designed, implemented, and evaluated the Complex Care program in this study were employed by Evolent Health, Inc. The authors are grateful for the contributions of the entire team.

Author Affiliations: Evolent Health (NO, OCT, SP, AC, YA), Arlington, VA.

Source of Funding: Evolent Health.

Author Disclosures: Mr O'Hara is employed by and owns stock in Evolent Health. Ms Tran is a past employee of Evolent Health. Mr Phatakwala is a past employee of and owns stock in Evolent Health. Dr Cattrell and Mr Ajami are employed by Evolent Health.

Authorship Information: Concept and design (OCT, SP, AC); acquisition of data (OCT, AC, YA); analysis and interpretation of data (NO, OCT, SP, YA); drafting of the manuscript (NO, OCT); critical revision of the manuscript for important intellectual content (NO, OCT, SP, AC); statistical analysis (NO, OCT, SP, YA); obtaining funding (AC); administrative, technical, or logistic support (NO, SP); and supervision (SP, AC).

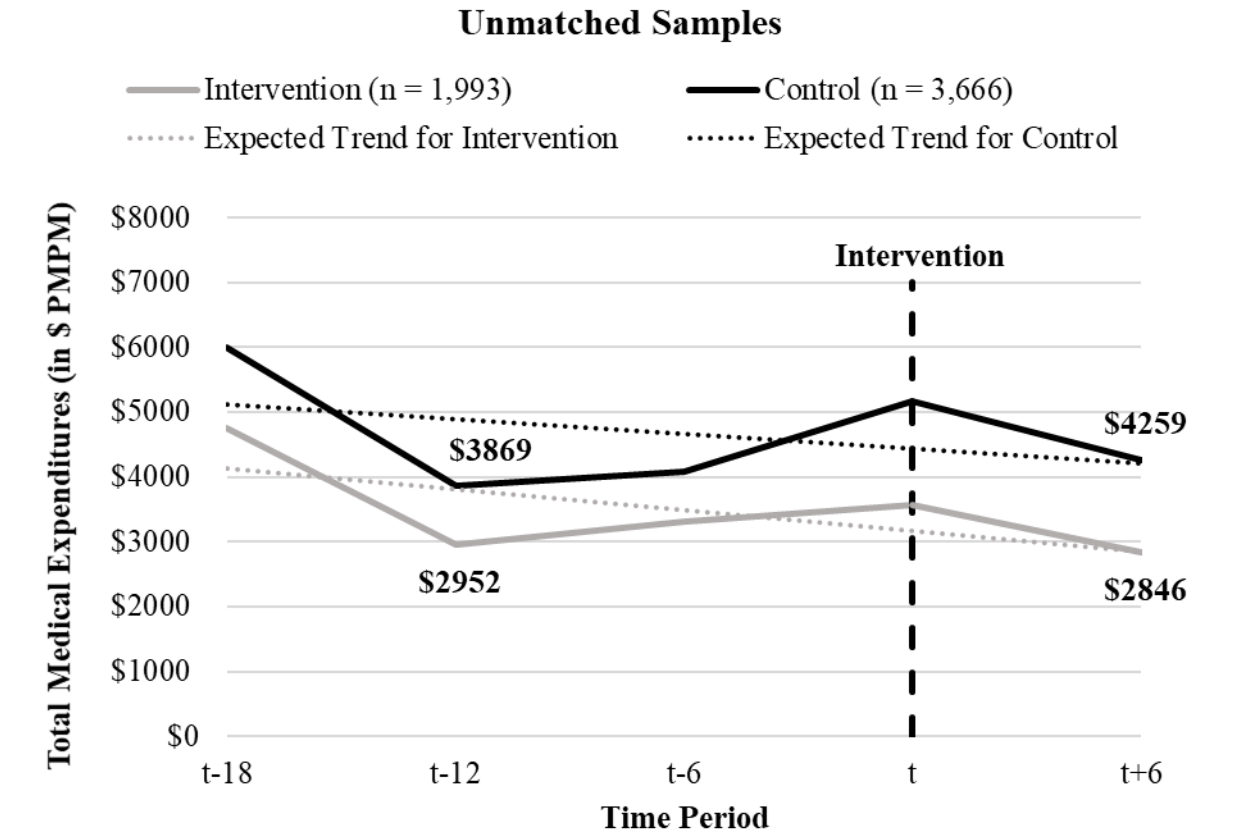
Address Correspondence to: Neal O'Hara, BA, Evolent Health, 800 N Glebe Rd, Arlington, VA 22207. Email: nohara@evolenthealth.com.

REFERENCES

1. Program data. CMS. Updated March 10, 2020. Accessed October 1, 2019. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/program-data.html
2. CMS, HHS. Medicare Program; Medicare Shared Savings Program; accountable care organizations—Pathways to Success and extreme and uncontrollable circumstances policies for performance year 2017. final rules. *Fed Regist*. 2018;83(249):67816-68082.
3. Next Generation ACO model. CMS. Updated April 23, 2020. Accessed January 19, 2019. https://innovation.cms.gov/initiatives/next-generation-aco-model
4. Lowell KH. Next Generation accountable care organization (NGACO) model evaluation: first annual report. CMS. August 27, 2018. Accessed January 19, 2019. https://innovation.cms.gov/Files/reports/nextgenaco-firstannrpt.pdf
5. Baker JM, Grant RW, Gopalan A. A systematic review of care management interventions targeting multimorbidity and high care utilization. *BMC Health Serv Res*. 2018;18(1):65. doi:10.1186/s12913-018-2881-8
6. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database Syst Rev*. 2016;3:CD006560. doi:10.1002/14651858.CD006560.pub3
7. Hsu J, Price M, Vogeli C, et al. Bending the spending curve by altering care delivery patterns: the role of care management within a Pioneer ACO. *Health Aff (Millwood)*. 2017;36(5):876-884. doi:10.1377/hlthaff.2016.0922
8. Peikes D, Chen A, Schore J, Brown R. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA*. 2018;301(6):603-618. doi:10.1001/jama.2009.126
9. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. *Health Aff (Millwood)*. 2012;31(6):1156-1166. doi:10.1377/hlthaff.2012.0393
10. McWilliams JM, Chernew ME, Landon BE. Medicare ACO program savings not tied to preventable hospitalizations or concentrated among high-risk patients. *Health Aff (Millwood)*. 2017;36(12):2085-2093. doi:10.1377/hlthaff.2017.0814
11. Sevak P, Stepanczuk CN, Bradley KWV, et al. Effects of a community-based care management model for super-utilizers. *Am J Manag Care*. 2018;24(11):e365-e370.
12. Gao J, Moran E, Li YF, Almenoff PL. Predicting potentially avoidable hospitalizations. *Med Care*. 2014;52(2):164-171. doi:10.1097/MLR.0000000000000041
13. Salzman BE, Knuth RV, Cunningham AT, LaNoue MD. Identifying older patients at high risk for emergency department visits and hospitalization. *Popul Health Manag*. 2019;22(5):394-398. doi:10.1089/pop.2018.0136
14. Hong CS, Siegel AL, Ferris TG. Caring for high-need, high-cost patients: what makes for a successful care management program? *Issue Brief (Commonw Fund)*. 2014;19:1-19. doi:10.15868/socialsector.25007
15. O'Malley AS, Peikes D, Wilson C, et al. Patients' perspectives of care management: a qualitative study. *Am J Manag Care*. 2017;23(11):684-689.
16. Pourmohammadi K, Hatam N, Shojaei P, Bastani P. A comprehensive map of the evidence on the performance evaluation indicators of public hospitals: a scoping study and best fit framework synthesis. *Cost Eff Resour Alloc*. 2018;16:64. doi:10.1186/s12962-018-0166-z
17. Hill B, Richardson B, Skouteris H. Do we know how to design effective health coaching interventions: a systematic review of the state of the literature. *Am J Health Promot*. 2015;29(5):e158-e168. doi:10.4278/ajhp.130510-lit-238
18. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. *Milbank Q*. 1996;74(4):511-544.
19. Lorig KR, Sobel DS, Stewart AL, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. *Med Care*. 1999;37(1):5-14. doi:10.1097/00005650-199901000-00003
20. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA*. 2002;288(19):2469-2475. doi:10.1001/jama.288.19.2469
21. Vermunt NPCA, Harmsen M, Westert GP, Olde Rikkert MGM, Faber MJ. Collaborative goal setting with elderly patients with chronic disease or multimorbidity: a systematic review. *BMC Geriatr*. 2017;17(1):167. doi:10.1186/s12877-017-0534-0
22. Coleman EA, Smith JD, Raha D, Min S-J. Posthospital medication discrepancies: prevalence and contributing factors. *Arch Intern Med*. 2005;165(16):1842-1847. doi:10.1001/archinte.165.16.1842
23. Yan J, Linn KA, Powers BW, et al. Applying machine learning algorithms to segment high-cost patient populations. *J Gen Intern Med*. 2019;34(2):211-217. doi:10.1007/s11606-018-4760-8
24. Powers BW, Yan J, Zhu J, et al. Subgroups of high-cost Medicare Advantage patients: an observational study. *J Gen Intern Med*. 2019;34(2):218-225. doi:10.1007/s11606-018-4759-1
25. Sieck CJ, Walker DM, Retchin D, McAlearney AS. The patient engagement capacity model: what factors determine a patient's ability to engage? *NEJM* Catalyst. March 13, 2019. Accessed July 17, 2019. https://catalyst.nejm.org/doi/full/10.1056/CAT.19.0001
26. Daw JR, Hatfield LA. Matching and regression to the mean in difference-in-differences analysis. *Health Serv Res*. 2018;53(6):4138-4156. doi:10.1111/1475-6773.12993
27. Ryan AM. Well-balanced or too matchy-matchy? the controversy over matching in difference-in-differences. *Health Serv Res*. 2018;53(6):4106-4110. doi:10.1111/1475-6773.13015
28. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Matching. In: *Impact Evaluation in Practice, Second Edition*. The World Bank; 2016:143-158. doi:10.1596/978-1-4648-0779-4_ch8
29. Adelson JL, McCoach DB, Rogers HJ, Adelson JA, Sauer TM. Developing and applying the propensity score to make causal inferences: variable selection and stratification. *Front Psychol*. 2017;8:1413. doi:10.3389/fpsyg.2017.01413
30. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994;47(11):1245-1251. doi:10.1016/0895-4356(94)90129-5
31. Difference-in-difference estimation. Columbia Mailman School of Public Health. Accessed October 8, 2019. https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation
32. Pérez D, Van der Stuyft P, Zabala MC, Castro M, Lefèvre P. A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. *Implement Sci*. 2016;11(1):91. doi:10.1186/s13012-016-0457-8
33. Zhang Z, Castelló A. Principal components analysis in clinical studies. *Ann Transl Med*. 2017;5(17):351. doi:10.21037/atm.2017.07.12
34. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care*. 2013;19(9):725-732.
35. The state of predictive analytics in U.S. healthcare. Modern Healthcare. Accessed January 18, 2019. https://www.modernhealthcare.com/assets/pdf/CH105920628.PDF
36. Hibbard JH, Greene J, Sacks RM, Overton V, Parrotta C. Improving population health management strategies: identifying patients who are more likely to be users of avoidable costly care and those more likely to develop a new chronic disease. *Health Serv Res*. 2017;52(4):1297-1309. doi:10.1111/1475-6773.12545
37. Cohen DJ, Dorr DA, Knierim K, et al. Primary care practices' abilities and challenges in using electronic health record data for quality improvement. *Health Aff (Millwood)*. 2018;37(4):635-643. doi:10.1377/hlthaff.2017.1254
38. Dorr DA, Cohen DJ, Adler-Milstein J. Data-driven diffusion of innovations: successes and challenges in 3 large-scale innovative delivery models. *Health Aff (Millwood)*. 2018;37(2):257-265. doi:10.1377/hlthaff.2017.1133
39. Jonk Y, Lawson K, O'Connor H, et al. How effective is health coaching in reducing health services expenditures? *Med Care*. 2015;53(2):133-140. doi:10.1097/mlr.0000000000000287

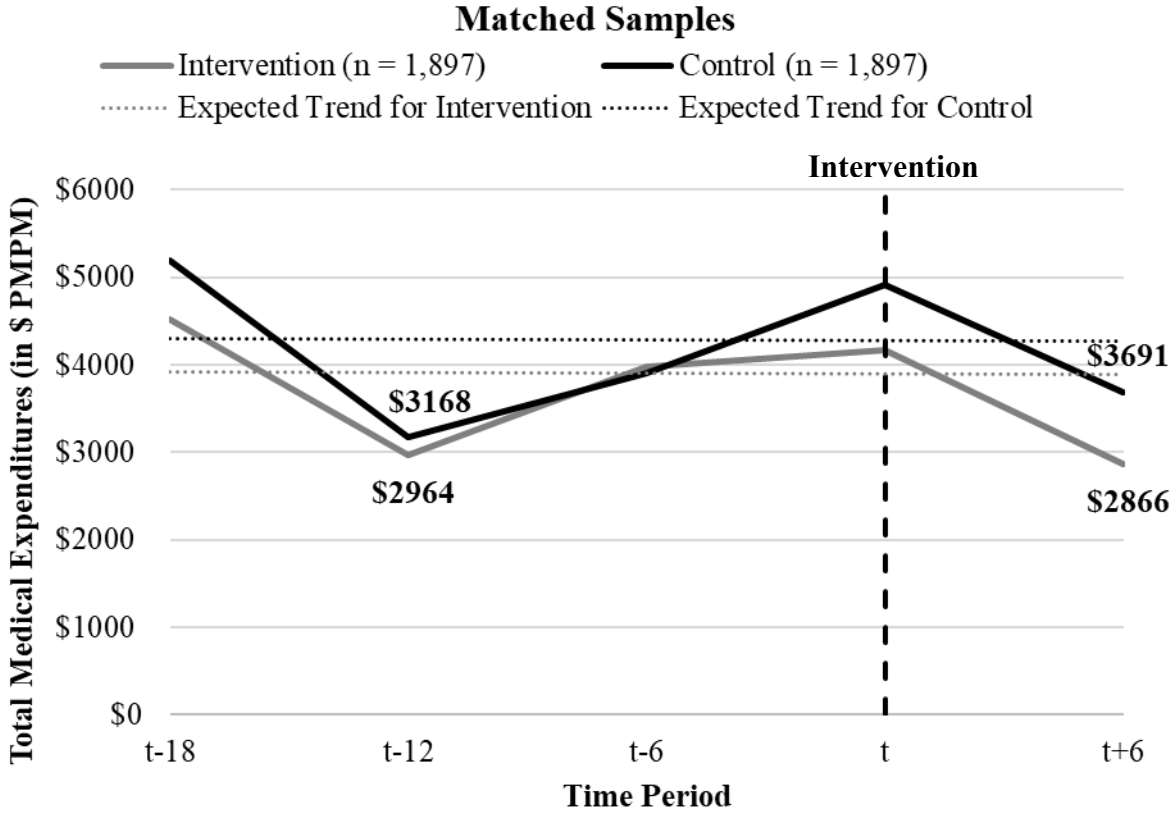
Visit **ajmc.com/link/4690** to download PDF and eAppendix

eAppendix A. Pre-Period Trends for Total Medical Expenditures (TME) per Member per Month (PMPM) for Unmatched Intervention and Control Samples



Each time period on the x-axis represents 6 months (e.g. *t-18* represents the period from 24 months to 18 months prior to the intervention). The expected trend was calculated using a linear regression line-of-best-fit based on the trend observed over 24 months in the pre-intervention time periods *t-18*, *t-12, t-6,* and *t*, which we then projected into the post-intervention period *t+6*.

eAppendix B. Pre-Period Trends for Total Medical Expenditures (TME) per Member per Month (PMPM) for Matched Intervention and Control Samples



Each time period on the x-axis represents 6 months (e.g. *t-18* represents the period from 24 months to 18 months prior to the intervention). The expected trend was calculated using a linear regression line-of-best-fit based on the trend observed over 24 months in the pre-intervention time periods *t-18*, *t-12, t-6,* and *t*, which we then projected into the post-intervention period *t+6*.

eAppendix C. Regression-Adjusted Propensity-Score-Matched Difference-in-Difference Results, by Control Sample Exclusion Criteria

A. Control Group includes patients who are: (1) in-queue for outreach; (2) contacted but not reached; and (3) contacted and declined to participate						
Control n = 1,897 *Intervention* n = 1,897	**Baseline[a]**		**Follow-Up[b]**		**Difference-in-Difference[c]**	
	Control	**Intervention**	**Control**	**Intervention**	**Percentage points (95% CI)**	***P***
Total Inpatient Admissions	1979.3	1854.2	2295.3	1743.6	-21.2 (-37.1 to -5.4)	0.03
ED Visits	1290.7	1294.2	1319.6	1288.2	-3.0 (-20.1 to 14.2)	0.78
Total Medical Expenditures	3167.7	2963.9	3691.2	2866.1	-22.0 (-37.6 to -6.5)	0.02
B. Control Group includes patients who are: (1) in-queue for outreach; and (2) contacted but not reached						
Control n = 946 *Intervention* n = 946	**Baseline[a]**		**Follow-Up[b]**		**Difference-in-Difference[c]**	
	Control	**Intervention**	**Control**	**Intervention**	**Percentage points (95% CI)**	***P***
Total Inpatient Admissions	2308	2056	2899	2069	-23 (-43 to -2)	0.07
ED Visits	1524	1436	1527	1438	0 (-21 to 21)	0.99
Total Medical Expenditures	3859	3378	4622	3310	-22 (-41 to -3)	0.06
C. Control Group includes patients who are: (1) in-queue for outreach						
Control n = 491 *Intervention* n = 491	**Baseline[a]**		**Follow-Up[b]**		**Difference-in-Difference[c]**	
	Control	**Intervention**	**Control**	**Intervention**	**Percentage points (95% CI)**	***P***
Total Inpatient Admissions	2859	2298	3465	2088	-21 (-43 to 0)	0.10
ED Visits	1657	1536	1574	1560	6 (-20 to 32)	0.71
Total Medical Expenditures	4369	3580	5122	3369	-17 (-36 to 2)	0.15

[a] Defined as 12 months prior to program enrollment or in queue status

[b] Defined as 6 months following program enrollment or in queue status

[c] Represents the relative change in each outcome for intervention group patients; estimates are from linear probability models adjusting for month of year and ACO program year as fixed effects and ACO as an indicator variable

eAppendix D. Variables Used in Propensity Score Model

Demographics

- Age
- Sex

Clinical Risk Assessments

- Charlson Comorbidity Score
- Predictive Risk Score Percentile

Cost

- Total Medical Expenditures (TME) (in $ PMPM)

Comorbid Diseases

- Diabetes (%)
- Asthma (%)
- Coronary Artery Disease (%)
- Chronic Obstructive Pulmonary Disorder (%)
- Congestive Heart Failure (%)
- Dementia (%)

Utilization Metrics (per 1000 members per month)

- Total Inpatient Admissions
  - Acute Admissions
  - Non-Acute Admissions
- ACSC Admissions
- ED Visits
- PCP Visits
- Specialist Visits