Evaluation of a High-Risk Case Management Pilot Program for Medicare Beneficiaries with Medigap Coverage

Kevin Hawkins, PhD,[1] Paula M. Parker, RN, CCP,[2] Cynthia E. Hommer, MSW, LICSW,[3]
Gandhi R. Bhattarai, PhD,[1] Jinghua Huang, MS, PhD,[1] Timothy S. Wells, MPH, PhD,[1]
Ronald J. Ozminkowski, PhD,[4] and Charlotte S. Yeh, MD[5]

Abstract

The objective was to evaluate the 3-year experience of a high-risk case management (HRCM) pilot program for adults with an AARP Medicare Supplement (Medigap) Insurance Plan. Participants were provided in-person visits as well as telephonic and mailed services to improve care coordination from December 1, 2008, to December 31, 2011. Included were adults who had an AARP Medigap Insurance Plan, resided in 1 of 5 pilot states, and had a Hierarchical Condition Category score > 3.74, or were referred into the program. Propensity score weighting was used to adjust for case-mix differences among 2015 participants and 7626 qualified but nonparticipating individuals. Participants were in the program an average of 15.4 months. After weighting, multiple regression analyses were used to estimate differences in quality of care and health care expenditures between participants and nonparticipants. Increased duration in the program was associated with fewer hospital readmissions. Additionally, participants were significantly more likely to have recurring office visits and recommended laboratory tests. The program demonstrated $7.7 million in savings over the 3 years, resulting in a return on investment of $1.40 saved for every dollar spent on the program. Savings increased each year from 2009 to 2011 and with longer length of engagement. The majority of savings were realized by the federal Medicare program. This study focused on quality of care and savings for an HRCM program designed solely for Medicare members with Medicare Supplement coverage. This program had a favorable impact on quality of care and demonstrated savings over a 3-year period. (*Population Health Management* 2015;18:93–103)

Introduction

CURRENTLY, MORE THAN 40 million Americans are at least 65 years of age, and this group has increased by approximately 15% since 2000.[1] By 2020, 55 million Americans are expected to be at least 65 years of age,[2] totaling 16% of the population. Advanced age is accompanied by an increased likelihood of multiple chronic conditions. For example, in 2010, two thirds, or about 21 million Medicare beneficiaries, had 2 or more chronic conditions, and 37% had 4 or more chronic conditions.[3] In addition, the 14% of beneficiaries with 6 or more chronic conditions accounted for 55% of Medicare spending on hospitalizations and 63% of expenditures on post-acute care, such as skilled nursing facilities and long-term care.[3] This group also was readmitted about 30% more often than those who were younger than 65 years of age.

Medicare beneficiaries with multiple chronic conditions often receive health care that is fragmented. A recent study found that most Medicare beneficiaries see 2 primary care physicians and 5 specialists working in 4 different practices.[4] The Institute of Medicine noted these shortcomings and recommended increasing the use of programs that can help Medicare beneficiaries better coordinate their care.[5] Similar concerns have been voiced by the Department of Health and Human Services.[6] Poor coordination of care can lead to higher health care expenditures. Increased spending on chronic conditions is one of the key factors related to the rapidly growing expenditures in the federal Medicare program,[7] making it prudent to reduce unnecessary spending when possible.

Case management is a comprehensive process designed to meet the health care needs of individuals with multiple

---

[1]OptumInsight, Ann Arbor, Michigan.
[2]Evolent Health, Arlington, Virginia.
[3]MCG—Formerly Milliman Care Guidelines, Seattle, Washington.
[4]Consumer Solutions Group, Optum, Ann Arbor, Michigan.
[5]AARP Services, Inc., Washington, DC.

chronic conditions, with the goal of ensuring that high-quality, cost-effective health care services are used.[8] To date, case management has been shown to improve health outcomes. For example, one study reported that individuals with 3 or more restrictions in activities of daily living and who had enrolled in a case management program had decreased nursing home admissions.[9] Similarly, other studies have reported positive findings associated with case management, such as slowed declines in health status,[10] lowered rehospitalization rates,[11,12] and improved self-reported social functioning and satisfaction.[13]

Another goal of case management is to reduce unnecessary costs, but few sound economic evaluations have been conducted and those performed have shown inconsistent results. A study by Duke focused on a sample of 175 participants who were at least 65 years of age.[14] In this study, emergency department and hospital admission expenditures were 36% and 60% lower, respectively, among those who were managed. Picariello et al conducted an evaluation of a case management program offered to Medicare Advantage beneficiaries and reported net savings of almost $5000 per participant per year.[15] Conversely, a 1-year randomized trial found no difference in health care costs associated with a case management program,[16] nor did an 18-month randomized controlled trial of 8504 beneficiaries enrolled in a Medicare Plus Choice health maintenance organization plan.[13] Finally, Baker et al conducted a study of a care coordination program for Medicare insureds and found the program resulted in about a 10% reduction of health care expenditures.[17] This study may have underestimated program impact by including a large proportion of people in the intervention group who did not actively engage in the intervention.

The present study focuses on Medicare fee-for-service beneficiaries who have a Medigap plan that carries the AARP brand. Currently, about 3.5 million adults have an AARP Medicare Supplement Insurance Plan. These plans are insured either by UnitedHealthcare Insurance Company or UnitedHealthcare Insurance Company of New York, and are offered in all 50 states, Washington, DC, and various US territories.

A high-risk case management (HRCM) program was offered to Medigap insureds who lived in target markets located in 5 states, including parts of California, Florida, New York, North Carolina, and Ohio. These states were not randomly selected. They were selected because there were a large percentage of AARP members who qualified for these programs and had UnitedHealthcare Medicare Part D coverage.

The HRCM program evaluated herein began in December 2008 and was ongoing through 2011. The first objective of this evaluation was to determine whether those who participated in this program received better quality of care than those who did not participate. Another key objective was to determine whether participation was associated with reductions in health care expenditures, yielding savings that might offset program costs and produce a positive return on investment (ROI) associated with program participation.

Methods

The HRCM program intervention

The HRCM program was voluntary and offered at no additional expense to qualified Medigap insureds in the 5 pilot states. The program used several sources to find individuals qualified for the program. Included were those who had Hierarchical Condition Category (HCC) scores greater than 3.74. This cutoff can be lowered if predictive modeling is used to identify those who are most likely to succeed in the program.

The HCC score is obtained from a standard risk adjustment tool utilized by the Centers for Medicare & Medicaid Services (CMS). The underlying HCC model developed by CMS uses a sample of over 1 million insureds to predict health care costs. These predicted costs are then converted to relative risk factors.[18] Thus, the average Medicare beneficiary has an HCC score of 1.0, while scores greater or less than 1.0 reflect expectations for higher or lower costs in the future based on the individual's age, sex, and medical conditions. HCC scores are refreshed monthly through an internal process for all AARP Medigap insureds and are available in the claims database.

Insured members also were invited into the program if their HCC scores were less than 3.74, but they were referred to the program after talking to a nurse for another reason. For example, if a member called the Nurse HealthLine program (which is available to all AARP members with a United Healthcare Medicare Supplement Insurance policy) to discuss another matter and it became clear during that conversation that the caller may qualify for the HRCM program, the caller would be referred to the HRCM program. Other insureds may have been referred to the HRCM program by their doctor or another care provider who knew of the program. Still others may have been referred after completing a health risk assessment survey and reporting process that yielded information about their health status suggesting that the HRCM program may be beneficial to them.

If an individual agreed to participate in the HRCM program he or she received an in-home visit by a nurse case coordinator. During that visit, a comprehensive assessment of the individual's health history was conducted. An advanced plan of care was developed and shared with the participant, his or her physician, and with any other caregivers, if the participant agreed to share this information with those people. Participants were then called on the telephone approximately every 3 weeks to discuss their plans of care and ongoing health status. In addition, if the participant was hospitalized, the nurse case coordinator assisted the hospital with discharge planning as well as planning for in-home care, if needed. Further, the nurse case coordinator was available to help participating members prepare for upcoming physician appointments or to accompany participants to their appointments if desired. Participants also received regular mailings regarding general health topics (eg, diet, exercise, immunizations). These mailings also included tailored messages regarding gaps in care, such as pharmaceutical refill reminders and missed office visits.

Study design

The HRCM program evaluation compared 2 types of individuals, those who participated in the program (program participants) and those who were qualified but did not participate in it (program nonparticipants). First, an overall evaluation was conducted that included those who participated in the program for at least 30 consecutive days during

the evaluation period, which ran from the program's inception in December 2008 through December 2011. Additional analyses tested if duration in the program was associated with better quality of care and health care expenditure outcomes. This was done by dividing participating members into 3 equal-sized groups based on the duration of program participation throughout the study period (ie, 1–9 months, 10–18 months, 19–37 months).

Three comparison groups of nonparticipants were randomly selected from among qualified nonparticipants so that outcomes observed for program participants could be compared to suitable reference groups before making inferences about program impact. One comparison group was observed for 1–9 months after index, another was observed for 10–18 months after index, and the third was observed for 19–37 months. Any participants or nonparticipants who had incomplete demographic information, who had less than 3 months of pre- and postintervention period insurance membership data, or who had negative health care expenditures were excluded from the study.

Index date assignment and associated time lags from program qualification

The evaluation design required that program participants be followed for periods of time before and after exposure to the program. In order to make comparisons, nonparticipants also had to be followed for a similar time period even though they had no exposure to the program. This design allowed the calculation of trends over time in health care expenditures and other outcomes for both groups of sample members. Comparing the trends for participants to trends for nonparticipants allowed for more accurate inferences of program impact, once statistical adjustments were applied to ensure these 2 groups of people were similar in terms of many other factors that also influenced outcomes.

To understand how health care expenditures changed over time for both participants and nonparticipants, index dates were defined for each individual to divide his or her observed time frames into 2 periods (before versus after index). The index date for participants was the date the individual joined the program. The index date for each nonparticipant was based on the date he or she became eligible for the program plus a time lag adjustment.

In any program there may be a time lag between becoming eligible for and engaging in a program. This is because it takes time to contact individuals, describe the advantages of the program, obtain a decision to participate, and then to set up and apply program services. The time lag for each engaged individual (ie, the time between the first date of program qualification and the date of actual program engagement) was determined and the distribution of these dates was applied to nonparticipants via a random assignment method to ensure the time lag distribution for each group was the same. The 6 months prior to the index date was then defined as the preintervention period, while the 12 months after the index date was the postintervention period. This procedure is illustrated in Figure 1.

Covariates

Program participants and nonparticipants may differ with regard to demographics, health status, and other factors. Thus the statistical analyses that were used to estimate the impact of program participation had to account for as many of these differences as possible. First, the demographic, health status, and other covariates used in these analyses are described.

Several demographic measures were accounted for in the evaluation of the HRCM program because these may independently influence health care utilization and expenditures. These demographics included the insured member's age, sex, and 2 variables measuring member location. The first location variable indicated if the individual resided in a rural versus an urban area, as defined using guidelines published by the US Census Bureau.[20] The second indicated the state of residence for the individual.

Socioeconomic variables also were included in the analyses because these may influence health care utilization and expenditures as well. These included zip code–level correlates of the individual's race and income, which were unknown. Ascribing labels to individuals based on the zip codes where they lived is obviously imperfect, but the research team believes it is better to at least try to account for socioeconomic factors in an imperfect way rather than make
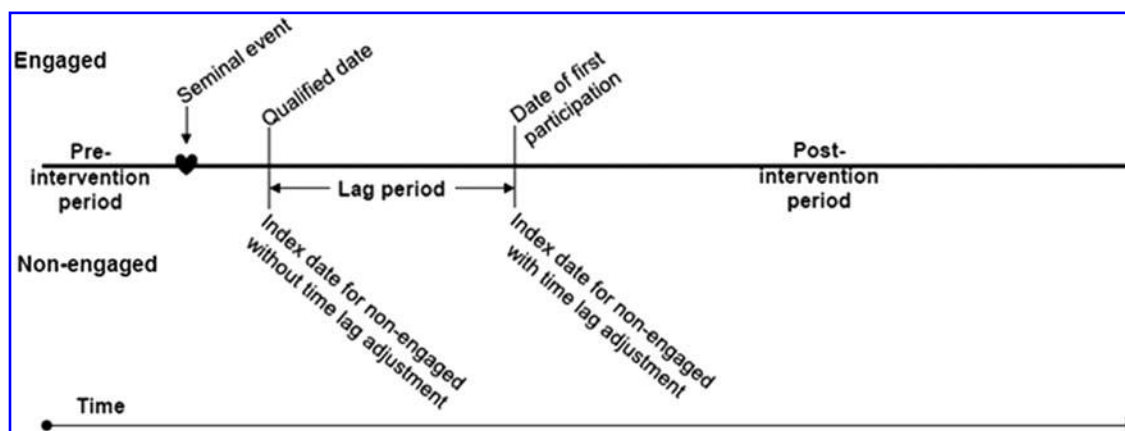

FIG. 1. Schematic describing how the index date for nonparticipants is established based on the lag in time from program eligibility to program participation among participants.

no adjustments at all for socioeconomic factors. Therefore, zip code–level correlates of race and income were used. Likely membership in various race or income groups has been shown to be correlated to zip code–based metrics, and these metrics have been shown to influence health care utilization and expenditures.[21]

Thus, in this study, the research team coded for whether the respondent lived in an area with more or fewer minority members, depending on the percent of minority residents living in the individual's zip code. This was accomplished by extracting 2010 US Census Bureau demographic data from the American FactFinder Web site.[22] The data obtained included population totals and a breakdown based on race, which included the categories of white, black, American Indian, Asian, Pacific Islanders, and Other. A minority status variable was then created using the ratio of nonwhite population to the total population, and binary indicators based on this ratio were used to account for the impact of living in low, medium, or high minority areas. Similarly, residence in areas of higher or lower income was coded as high, medium-high, medium, or low, based on whether the median income in the individual's zip code area was in the highest, second-highest, third-highest, or lowest quartile in 2010, according to US Census records.

Next, the type of Medigap insurance plan was accounted for because insurance plan features may influence utilization of services and health care expenditures. Members who had more complete coverage for Medicare deductibles and co-insurance were those who enrolled in Medicare supplement plans C, F, or J. An indicator was created to account for this in the analysis.

The research team also accounted for whether sample members participated in other health care coordination programs that were offered in the pilot states. These other programs include the aforementioned Nurse HealthLine program, pharmaceutical adherence or medication therapy management programs, and disease management programs for congestive heart failure (CHF), coronary artery disease (CAD), diabetes, or depression. The team also adjusted for the calendar year in which HRCM program qualification was first observed to account for the likelihood that program maturation may be associated with better outcomes.

Several health status measures also were included in the analyses, as these may influence health care utilization and expenditure trends. First among these was the HCC score. Also included were indicators for whether diagnoses of dementia or psychoses were evident in the individual's medical claims data. Dementia and psychosis are conditions that are associated with long-term memory loss and may be challenging and very expensive to manage. Therefore, indicators for having either of these 2 diagnoses are included in the participation and cost-saving models. Models also are adjusted for whether there was evidence of placement in long-term nursing facilities during the preintervention period.

The research team also accounted for whether individuals had emergency room visits or hospitalizations in the 6 months prior to their index date. In addition to influencing subsequent utilization and expenditure trends, these variables may help account for differences in motivation to care for oneself.

Next, pharmaceutical claims data were available for approximately half of the sample members, namely those who had Medicare Part D coverage provided by UnitedHealthcare Insurance Company. The primary goal of Medicare Part D is to ensure that Medicare beneficiaries will not forgo necessary prescription drugs because they cannot afford them.[23] Therefore, differences in Medicare Part D coverage among participants and nonparticipants may influence the expenditures being evaluated. A binary flag indicated whether or not a member had prescription drug coverage.

The analyses also accounted for differences in the supply of health care services in the area where the individual lived, as these are known to influence health care utilization and expenditures.[24] These variables were derived from the Dartmouth Atlas of Healthcare.[25] These data are at the hospital service area (HSA) level,[26] which is then transferred to zip code–level by the crosswalk supplied with the data. Physician-based metrics were calculated per 100,000, while the number of hospital beds was calculated per 1000 residents in the individual's HSA.

Study outcomes used to assess program impact

All of the aforementioned variables were used in statistical analyses to help estimate the impact of program participation on quality of care and health care expenditures. The quality metrics are described next, followed by a description of how expenditure trends were used to make inferences about whether the HRCM program saved money for participants and the Medicare program.

Quality metrics. To determine if the HRCM program was associated with improvements in quality of care, compliance with a number of evidence-based guideline (EBG) metrics was noted. EBGs are treatments that are consistent with processes described in the peer-reviewed literature that are associated with significantly higher quality of care.[27]

Quality of care was defined based on the percentage of individuals with a quality gap in the preintervention period who closed that gap in the postintervention period by receiving services that are associated with higher quality of care. Quality gaps were defined as a lack of evidence in the claims data that an EBG metric was met. Some of these quality gaps were not disease specific, while others were related to particular conditions that were prevalent among those who qualified for the HRCM program.

Non-disease-specific EBG metrics included a metric indicating whether a sleep agent was used after an accidental fall or hip fracture. Other non-disease-specific metrics were indicators of whether drugs were prescribed that should be avoided in the elderly. These drugs are noted on the Beers list, which contains drugs that meet specific criteria as being potentially inappropriate medications in older adults.[28] An example of a Beers-listed drug is a tricyclic antidepressant or anticholinergic agent prescribed to someone with previous claims evidence of a diagnosis consistent with dementia.

Other EBG metrics were related to heart disease, diabetes, or CHF. These disease-specific EBG metrics included having a low-density lipoprotein (LDL) cholesterol test in the last 12 reported months, compliance with prescribed beta-blockers or angiotensin-converting enzyme inhibitors, having at least 2 hemoglobin A1c (HbA1c) tests in last 12 reported months among patients with diabetes, or having 1 or more office visits to manage either CAD, diabetes, or

CHF in the last 12 reported months before the member's study end date.

In addition to the EBG metrics, other quality of care metrics were based on having readmissions for any cause and to any hospital within 30 days of a previous hospital discharge, and whether or not the individual had an office visit with his or her health care provider within 15 days of hospital discharge. Readmissions that occur shortly after previous discharge are often caused by inadequately resolved issues from the prior hospitalization.[29] Additionally, hospital readmissions can be caused by deterioration in the individual's health shortly after discharge, which may have been avoided if the individual had received appropriate care after discharge.[29]

ROI. The ROI in the HRCM program was evaluated first by estimating savings in health care expenditures associated with program participation, then by comparing those savings to the cost of operating the program. Savings that exceed program costs provide evidence of a positive ROI for the program.

Savings were estimated using a difference-in-difference approach. Specifically, average monthly health care expenditures in the period before program participation was offered were calculated, as were average monthly expenditures after it was offered. The differences in these averages over time were compared for program participants and qualified nonparticipants. Savings or losses were inferred depending on whether changes over time in health care expenditures per month were higher for program participants or nonparticipants. The aforementioned propensity score and multiple regression analyses accounted for differences in the covariates mentioned earlier, before savings were estimated. The regression process is described in more detail later in this paper. Health care expenditures included payments from Medicare, the Medigap Plan, and the member's out-of-pocket contribution.

Once savings were estimated, they were compared to HRCM program operating costs in order to estimate an ROI ratio for the program. This ratio was calculated as the total savings associated with program participation divided by total costs associated with participation in it. An ROI ratio greater than 1.0 would indicate net savings for the program, while an ROI ratio less than 1.0 would indicate net losses associated with program participation.

Statistical analyses used to estimate program savings and quality of care

Statistical analyses were conducted by combining the data from each duration of participation period group to yield an overall (3-year) program impact estimate. Three sets of analyses were conducted. The first set of analyses included descriptive statistics that described the demographic and other characteristics of the sample and compared the unadjusted case mix differences between participants and nonparticipants. Chi-square and Student *t* tests were used to test for differences in categorical and continuous variables, respectively. Secondly, propensity score weighting, as described in the following section, was used to remove most of the case-mix differences between participants and nonparticipants. The third set of analyses used additional multiple

regressions as a further control for demographic, socioeconomic, location, and health status differences before making inferences about program impact on quality of care or health care expenditure trends. This helped to account for any remaining case-mix differences between participants and nonparticipants after propensity score weighting, and adjusted for skewed distributions of health care expenditures, which are common.[30] Additionally, previous research has shown that increased program duration is associated with better outcomes,[31] and evaluations were completed for each duration period to determine how different lengths of program participation affected study findings.

Propensity score weighting. Propensity score weighting[32,33] was used to minimize the case-mix differences between participants and nonparticipants. To perform the propensity score weighting, a logistic regression model was used to estimate the likelihood of participating in the program. The variables used in the logistic model were the covariates previously described. The propensity score for each sample individual was then obtained from the results of the logistic regression, and was defined as his or her predicted probability of being in the HRCM program. This probability was then used to construct a weight to be used in subsequent analyses to adjust for case-mix differences. The weights were defined as inverse probability weights, which were calculated as 1/predicted probability of HRCM participation for each participant, and as 1/(1-predicted probability of HRCM participation) for each nonparticipant. As suggested in the literature, weights were normalized to equal the sample sizes in each group.[34] Applying the weights when comparing the mean values for demographics, health status, and outcomes helped to adjust for case-mix differences between those in the participant and nonparticipant groups. The literature has shown that propensity score weighting is a convenient and acceptable way to remove case-mix differences when evaluating health and wellness programs.[35]

Analyses for ROI and quality estimates. Per member per month (PMPM) savings were measured via multiple regressions that applied a difference-in-difference design, so the research team could learn whether trends in health care expenditures differed over time for program participants versus nonparticipants. The multiple regression approach utilized ordinary least squares regressions. The difference in the PMPM total health care expenditure in the pre-intervention versus postintervention periods was used as the dependent variable value for each sample member. Primary explanatory variables included whether or not the individual was engaged, an indicator for whether the outcome was measured before or after the index date, and an interaction term for these 2. Models also were adjusted for the demographic, health status, location, and the supply side variables mentioned previously.

The ROI estimates were generated from the savings regression models, as mentioned earlier, and were measured for all 3 years together (ie, an overall model) by participation duration group, and by calendar year.

To calculate the quality estimates, separate logistic models were performed for each previously described quality metric as the dependent variable, participation status (yes/no) was the primary explanatory variable, and the same independent variables that were used in the expenditure models were used

to adjust for case-mix differences. Odds ratios greater than 1.0 would suggest a benefit associated with participation with regard to the occurrence of office visits within 15 days of previous discharge, having an LDL or HbA1c blood test, the use of beta-blocker drugs for those who need them, nonuse of drugs to be avoided by the elderly, and having an annual office visit. Conversely, odds ratios less than 1.0 would suggest a benefit associated with participation related to avoided hospital readmissions.

Sensitivity analyses. To determine if the results were sensitive to outliers, propensity weighted models with and without outliers were estimated to gauge if the exclusion of a few individuals with either very high or very low expenditures influenced the results. The initial analyses were conducted after excluding outliers. In the sensitivity analysis, outliers were added back into the sample. Outliers were identified using a method first described by Heckman et al.[36] The intent of the outlier identification method was to ensure that the ranges of health care expenditures from lowest to highest were identical for HRCM participants and nonparticipants. Expenditures outside the common range in each group were labeled as outliers and removed from the main analyses but included in the sensitivity analyses.

An additional sensitivity analysis was conducted by estimating savings using propensity score matching. Propensity score matching[32,37] was accomplished by using the propensity scores to conduct a one-to-one matching of participants to nonparticipants using a reasonable minimum difference, or caliper size, based on the predicted probability score. A within caliper, nearest neighbor matching technique was used to match participants and nonparticipants based on their propensity scores. Finally, models were run without the race and income variables that were based on zip code–level data to determine if results were sensitive to the inclusion or exclusion of these variables.

Results

This evaluation included 2015 participants and 7626 nonparticipants who were qualified to participate but chose not to do so. Separate analyses were conducted by time spent engaged in the program, as previously described. Within each participation duration group, prior to propensity score weighting, there were significant differences in a number of the means or percentages among the case mix measures between participants and nonparticipants. After propensity score weighting, most measured case-mix differences were removed. This allowed program impact to be estimated more accurately for each program duration group, as case-mix differences no longer influenced the results (Table 1).

In the overall model, there were 2015 participants in the propensity score weighted analysis that had been in the program for at least 30 consecutive days during December 2008 through December 2011. Participants had an average of 15.4 months of time in the postintervention period. The PMPM savings were $250, resulting in an overall program savings of $7.7 million, and the overall ROI was 1.4:1, indicating that for every dollar invested in the program, the program saved $1.40. The ROI estimates varied with duration in the program and by year of engagement, although

only 2 of these differences were statistically significant (Table 2). The program savings increased as the year of engagement increased. From 2009 to 2011 the ROI increased each year (from −0.3:1 to 3.9:1). Additionally, shorter program engagement (1–9 months) was associated with a lower ROI compared to program engagements of 10–18 months or 19–37 months (Table 2).

The results from the sensitivity analysis suggest that the ROI estimates were sensitive to the inclusion or exclusion of a few individuals with very high or low expenditures. With the exception of the 1–9 month group, ROI estimates decreased when outliers were included.

In contrast, the magnitude and significance levels of the ROIs were not sensitive to the inclusion or exclusion of the zip code–level race and income variables. The exclusion of these variables resulted in the ROI changing by a maximum of 0.4, where the ROI for the 1–9 month was −5.3:1 with these variables compared with −4.9:1 for the model without these variables. In addition, the ROI P value for those first engaged in 2010 went from statistically nonsignificant ($P = 0.055$) to statistically significant ($P = 0.041$). (Data not shown.)

Propensity score matching resulted in 1604 matches, and included 79.6% of the participants and 21% of the nonparticipants. Results were comparable to those obtained using propensity score weighting. In this analysis, participants had an average of 15.8 months of time in the postintervention period. The PMPM savings were $329, resulting in an overall program savings of $8.3 million, and the overall return on investment (ROI) was 1.8:1, which is slightly higher than the one calculated using propensity score weighting (1.4:1). The difference between the propensity score matched results compared with the propensity score weighted results is most likely attributable to the propensity matched analyses included only 79.6% of the participants and 21% of the nonparticipants who were included in the propensity weighted results. (Data not shown.)

In the analyses of the quality metrics, 13 of the 19 metrics tested (68.4%) favored participation, although most were not statistically significant (Table 3). Among the statistically significant findings, those in the 1–9 month group were more likely to be readmitted within 30 days of a previous admission, were less likely to have an office visit within 15 days of a discharge from a hospitalization, and were more likely to have an annual office visit to manage CAD, diabetes, or CHF. The only other statistically significant finding was that participants in the 19–37 month group were more likely to have an LDL test compared with nonparticipants.

Discussion

To the research team's knowledge, this study is a unique evaluation of quality outcomes and ROI associated with an HRCM program for Americans who were generally at least 65 years of age, and were Medicare beneficiaries with Medigap coverage. In this study, the preponderance of evidence suggests that this program led to improvements in quality of care. In addition, the program was associated with a total savings of $7.7 million and an overall ROI of 1.4:1. In addition, the ROI increased with increasing time spent in the program and with later year of initial program participation.

TABLE 1. STUDY CHARACTERISTICS OF PARTICIPANTS AND NONPARTICIPANTS AFTER PROPENSITY SCORE WEIGHTING

Characteristic	1–9 Month group		10–18 Month group		19–37 Month group	
	Participant n=619 % or mean	Nonparticipant n=3586 % or mean	Participant n=608 % or mean	Nonparticipant n=2509 % or mean	Participant n=788 % or mean	Nonparticipant n=1531 % or mean
Age						
≤74 years	20.6%	21.0%	21.6%	23.6%	23.3%	22.1%
75–84 years	41.9%	41.8%	39.7%	39.5%	43.4%	44.4%
≥85 years	37.5%	37.2%	38.7%	36.9%	33.3%	33.5%
Sex						
Female	51.3%	54.3%	51.3%	54.3%	59.6%	56.9%
Race						
High minority	8.4%	8.1%	7.4%	7.7%	7.2%	7.5%
Medium minority	34.4%	38.1%	39.5%	38.3%	41.0%	39.1%
Low minority	57.1%	53.8%	53.2%	54.1%	51.8%	53.4%
Income						
High	56.1%	57.1%	55.4%	57.5%	56.8%	54.8%
Upper Medium	4.6%	6.2%	5.3%	6.1%	6.6%	7.3%
Lower Medium	15.0%	12.3%	11.1%	13.0%	11.7%	13.3%
Low	23.3%	22.8%	27.4%	22.5%	23.6%	23.7%
Missing	1.1%	1.6%	0.9%	1.0%	1.2%	1.0%
State						
California	17.0%	16.7%	16.7%	16.1%	13.0%	14.1%
Florida	11.4%	11.2%	10.8%	11.9%	12.6%	13.2%
North Carolina	8.1%	9.1%	8.3%	8.3%	11.1%	10.7%
New York	43.4%	42.8%	41.6%	43.5%	45.2%	42.9%
Ohio	16.8%	18.5%	20.6%	18.6%	15.8%	16.3%
Other	3.3%	1.8%	2.1%	1.7%	2.4%	2.8%
Metropolitan residence	96.5%	96.4%	97.5%	96.8%	94.7%	95.9%
Medigap plan type						
Plan C, F, or J	68.5%	62.6%	62.5%	62.5%	63.3%	64.2%
Participation in other programs prior to index						
Multiple in pre period	8.2%	7.7%	8.9%	8.6%	9.8%	9.4%
Multiple in post period	9.1%	1.4%	12.5%	3.6%	24.1%	5.9%
Disease management or level 2 case management						
Pre period	5.1%	5.9%	7.8%	6.9%	6.4%	5.5%
Post period	2.4%	3.5%	4.4%	4.1%	5.5%	5.3%
Part D drug coverage[a]	58.4%	58.1%	59.0%	57.5%	59.9%	59.8%
HCC score						
<2.8	18.3%	19.2%	18.4%	18.8%	30.8%	28.8%
2.8–3.74	15.6%	16.0%	14.3%	14.2%	17.1%	16.7%
3.75–4.269	23.7%	25.5%	25.5%	27.7%	23.9%	23.9%
≥4.27	42.4%	39.3%	41.8%	39.3%	28.3%	30.6%
Diagnosed with dementia						
Pre period	13.2%	12.1%	9.7%	10.0%	8.1%	8.7%
Post period	11.5%	7.0%	8.7%	11.3%	13.0%	14.5%
Diagnosed with psychosis						
Pre period	31.2%	27.5%	26.9%	24.9%	21.1%	21.2%
Post period	24.1%	16.1%	25.2%	23.8%	29.8%	33.0%
Long-term nursing home placement						
Pre period	38.3%	38.8%	38.1%	36.2%	30.0%	29.5%
Post period	18.8%	22.9%	27.6%	27.4%	32.2%	38.5%
Pre period ER visits[b]	24.1%	22.3%	23.4%	23.1%	23.6%	24.0%
Post period ER visits[c]	25.0%	17.0%	22.0%	19.6%	21.0%	18.1%
Hospital admissions						
Pre period[b]	24.9%	27.0%	27.7%	29.3%	30.1%	30.5%
Post period[c]	49.0%	20.8%	27.4%	24.5%	26.5%	24.4%
Acute care hospital beds[d]						
<2.2	35.0%	30.7%	32.8%	31.7%	33.8%	32.5%
2.2–2.8	29.0%	34.5%	36.0%	35.1%	31.6%	34.6%
≥2.9	36.0%	34.8%	31.3%	33.3%	34.6%	32.9%

TABLE 1. (CONTINUED)

| Characteristic | 1–9 Month group | | 10–18 Month group | | 19–37 Month group | |
	Participant n=619 % or mean	Nonparticipant n=3586 % or mean	Participant n=608 % or mean	Nonparticipant n=2509 % or mean	Participant n=788 % or mean	Nonparticipant n=1531 % or mean
Primary care physicians[e]						
<70	29.7%	27.3%	26.7%	28.0%	30.7%	29.1%
70–99	32.7%	36.5%	37.1%	37.1%	33.7%	35.5%
≥100	37.7%	36.3%	36.2%	34.9%	35.6%	35.5%
Index year						
2009	38.7%	39.2%	39.7%	46.3%	61.5%	64.3%
2010	34.2%	30.6%	47.8%	41.4%	38.5%	35.7%
2011	27.1%	30.3%	12.5%	12.3%	n/a	n/a

HCC, Hierarchical Condition Category; ER, emergency room.
[a]Only Part D coverage provided by UnitedHealthcare.
[b]In the 6 months prior to index.
[c]In the first 6 months of the postintervention period.
[d]Distribution is based on the number of acute care hospital beds per 1000 residents.
[e]Distribution is based on the number of primary care physicians per 100,000 residents.

Although not statistically significant, results suggest that longer length of engagement in the HRCM program was associated with fewer hospital readmissions. Also, the odds of having physician office visits shortly after hospital discharge were higher among the participants and increased with longer duration of participation. In contrast, there were no trends by duration of participation for EBG quality metrics. These findings are roughly consistent with a meta-analysis of 12 studies of hospital-based case management programs that found a 6% decrease in the readmission rate for those who participated in these programs.[38] Also, improvements in quality of care measures among participants were observed in a study of a university-based medical management center.[39] Taken together, these findings suggest that implementing similar case management programs may improve some quality

TABLE 2. SAVINGS, COSTS, AND ROI BY YEAR OF FIRST PARTICIPATION AND PARTICIPATION LENGTH

Category	Savings[a]	P value	Costs	Return on investment[b]
By year of first participation				
2009	-$58	.789	$181[c]	-0.3:1
2010	$416	.055	$181[c]	2.3:1
2011	$710	.028	$181[c]	3.9:1
By duration of participation				
1–9 months	-$967	.002	$181[c]	-5.3:1
10–18 months	-$81	.727	$181[c]	-0.5:1
19–37 months	$216	.244	$181[c]	1.2:1
Overall	$250	.093	$181[c]	1.4:1

[a]Savings are per member per month (PMPM) and calculated using a second stage regression model after propensity score weighting, where negative savings indicate a loss.
[b]Return on investment (ROI) is the ratio of savings to costs, where ROIs greater than 1.0:1 indicate a savings for every dollar invested in the program and ROIs less than 1.0:1 indicate a loss for every dollar invested in the program. For example, the Overall ROI of 1.4:1 indicates that for every $1.00 invested in the program, the program saved $1.40.
[c]Costs were based upon a blended cost for all 3 years.

of care metrics, perhaps by improving care coordination, as suggested by the Institutes of Medicine.[40]

The ROI for the program appears to be improving over time. The ROI for those who began participation in 2009 was −0.3:1 compared to 3.9:1 for those who began participation in 2011. This difference is likely related to program improvements and greater efficiencies as the program matured over this period of time.

In this study, the ROI of the HRCM program increased with longer duration in the program. There are at least 2 plausible explanations for this finding. When compared with those who were qualified but chose not to participate, the percentage of participants who died in each duration group was highest for the 1–9 month group and lowest for the 19–37 month group, and the pattern for health care expenditures was similar (Fig. 2). These findings suggest that those who are in their final phases of life accrue significant health care expenditures, lowering program ROI for those who participated for fewer than 10 months. Additionally, it likely takes several months of participation to improve the quality of health care of those in the program, and improved quality of care is generally regarded as a prerequisite to reduced health care expenditures and program savings.

The HRCM program staff develop close relationships with program participants through frequent phone contacts and need-driven face-to-face visits occurring over several months. Visits provide a comprehensive understanding of patient and caregiver needs and environmental issues, including home safety. In addition to visits, frequent phone contact provides ongoing continual support. An evaluation of another care coordination program offering a similar holistic approach for individuals enrolled in a high-risk Medicare health maintenance organization reported positive savings of $6.60 PMPM.[41] It is likely that bending the cost curve in care coordination programs is most successful when the program establishes close relationships with participants and utilizes a holistic approach.

Study limitations include the use of medical claims data that were collected for insurance purposes rather than for research purposes. It is difficult to measure health status well

TABLE 3. ODDS RATIOS FOR QUALITY METRICS

Quality metric	1–9 Month group Odds ratio	10–18 Month group Odds ratio	19–37 Month group Odds ratio	Overall Odds ratio
Hospital readmission within 30 days of previous admission[a]	1.53*	1.16	0.98	1.11
Office visit within 15 days of previous admission[b]	0.66**	1.07	1.13	1.04
Drug to be avoided in the elderly[b]	1.07	1.13	0.95	1.17
LDL cholesterol test (CAD or diabetes)[b]	1.28	1.64	2.37*	1.37*
Use of beta-blockers (CAD or CHF)[b]	0.79	c	c	0.81
Office visits (CAD, diabetes or CHF)[b]	1.55*	1.39	1.72	1.58***
Hemoglobin A1c test (diabetes)[b]	1.33	1.39	0.97	1.19

LDL, low-density lipoprotein; CAD, coronary artery disease; CHF, congestive heart failure.

[a]Odds ratios less than 1.0 indicate an improvement in the quality metric for participation compared with nonparticipation. Conversely, odds ratios greater than 1.0 indicate no improvement in the quality metric for participation compared with nonparticipation.

[b]Odds ratios greater than 1.0 indicate an improvement in the quality metric for participation compared with nonparticipation. Conversely, odds ratios less than 1.0 indicate no improvement in the quality metric for participation compared with nonparticipation.

[c]Insufficient volume to conduct meaningful comparisons.

*$P < .05$; ** $P < .01$; *** $P < .001$.

with claims data so some unknown amount of measurement error may affect these results. Next, the limitation of using zip code–level data to measure socioeconomic factors has already been mentioned, and inferences for those variables should be made with reference to residence in these areas, regardless of the individual's socioeconomic characteristic of interest. However, zip code–level data have been successfully used in the past for similar research exercises.[42] Finally, as with all statistical models that use multivariable adjustment, the propensity score and regression approaches may not have fully adjusted for motivational or other differences between participants and nonparticipants that could not be measured but may have influenced the results.

The strengths of this study include its relatively large sample size, the use of propensity score weighting and multiple regression methods to adjust for case-mix differences between participants and nonparticipants, and the examination of outcomes associated with different lengths of program participation. Including a number of sensitivity analyses also allowed inferences to be drawn regarding the effect of those with very high or low expenditures and different analytic methodologies on study findings.

In conclusion, the HRCM program is unique among Medigap plans, and has the advantages of not requiring provider incentives or having a limited network of health care providers from which to choose. This model aligns with current health care models that emphasize a more prominent consumer engagement role. In this evaluation, the HRCM program was associated with improvements in quality of care and a reduction in health care expenditures that was more likely to offset program costs in more recent years and as duration in the program increased. This study described the quality improvements and ROI associated with an HRCM program among a Medigap population. These findings may be used to establish a baseline for future models designed to further improve health care for these insured members moving forward.
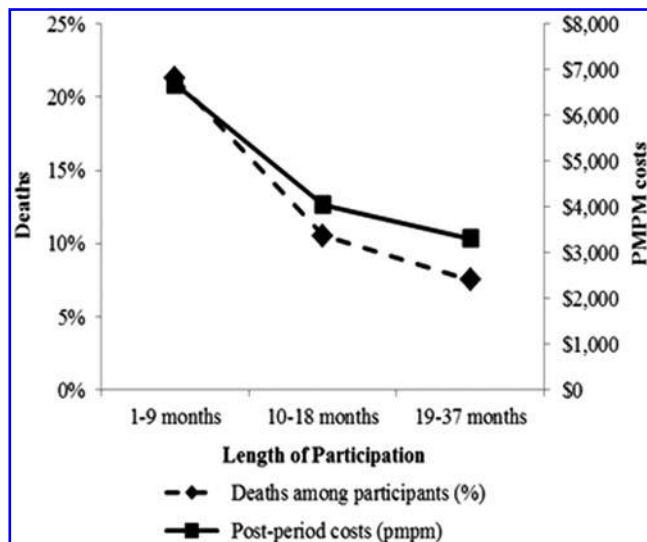
Acknowledgment

Author Disclosure Statement

FIG. 2. Mortality for high-risk case management program participants. Deaths and per member per month costs decrease with increased length of participation.

Health Insurance Program. The investigators retained full independence in the conduct of this research.

References

1. Howden LM, Meyer JA. Age and sex composition: 2010. US Census Bureau; 2011. http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf. Accessed November 15, 2011.

2. Administration on Aging. A profile of older Americans: 2011. US Department of Health and Human Services, Administration on Aging; 2011. http://www.aoa.gov/Aging_Statistics/Profile/2011/docs/2011profile.pdf. Accessed July 2, 2013.

3. Centers for Medicare & Medicaid Services. Chronic Conditions among Medicare Beneficiaries. Chartbook: 2012 Edition. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2012Chartbook.pdf. Accessed August 20, 2013.

4. Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. N Engl J Med. 2007;356:1130–1139.

5. Cruz M, Cruz RF. Compliance and costs in a case management model. Community Ment Health J 2001;37(1):69–77.

6. Rosenbaum PR. Model-based direct adjustment. Am Stat. 1987;82(398):387–394.

7. Riley GF. Long-term trends in the concentration of Medicare spending. Health Aff (Millwood). 2007;26:808–816.

8. American Case Management Association. Definition of case management; 2002. http://www.acmaweb.org/section.asp?sID=4&mn=mn1&sn=sn1&wpg=mh. Accessed January 28, 2014.

9. Newcomer R, Maravilla V, Faculjak P, Graves MT. Outcomes of preventive case management among high-risk elderly in three medical groups: a randomized clinical trial. Eval Health Prof. 2004;27:323–348.

10. Marshall BS, Long MJ, Voss J, Demma K, Skerl KP. Case management of the elderly in a health maintenance organization: the implications for program administration under managed care. J Healthc Manag. 1999;44:477–491.

11. Coleman EA, Parry C, Chalmers S, Min SJ. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med. 2006;166:1822–1828.

12. Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999;281:613–620.

13. Martin DC, Berger ML, Anstatt DT, et al. A randomized controlled open trial of population-based disease and case management in a Medicare Plus Choice health maintenance organization. Prev Chronic Dis. 2004;1(4):A05.

14. Duke C. The frail elderly community-based case management project. Geriatr Nurs. 2005;26:122–127.

15. Picariello G, Hanson C, Futterman R, Hill J, Anselm E. Impact of a geriatric case management program on health plan costs. Popul Health Manag. 2008;11:209–215.

16. Boult C, Rassen J, Rassen A, Moore RJ, Robison S. The effect of case management on the costs of health care for enrollees in Medicare Plus Choice plans: a randomized trial. J Am Geriatr Soc. 2000;48:996–1001.

17. Baker LC, Johnson SJ, Macaulay D, Birnbaum H. Integrated telehealth and care management program for Medicare beneficiaries with chronic disease linked to savings. Health Aff (Millwood). 2011;30:1689–1697.

18. Pope GC, Kautter J, Ingber MJ, Freeman S, Sekar R, Newhart C. Evaluation of the CMS-HCC Risk Adjustment Model. RTI International; 2011. http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf. Accessed April 8, 2014.

19. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999;282:1737–1744.

20. US Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria. http://www.census.gov/geo/reference/ua/urban-rural-2010.html. Accessed April 8, 2014.

21. American Hospital Association. Geographic variation in health care spending: a closer look. 2009. http://www.aha.org/research/reports/tw/twnov09geovariation.pdf. Accessed April 14, 2014.

22. US Census Bureau. American FactFinder. http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed April 15, 2014.

23. Munson JC, Morden NE, Goodman DC, Valle LF, Wennberg JE. The Dartmouth Atlas of Medicare Prescription Drug Use. 2013. http://www.dartmouthatlas.org/downloads/reports/Prescription_Drug_Atlas_101513.pdf. Accessed April 9, 2014.

24. Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. Health Aff (Millwood). 2002;Suppl Web Exclusives:W96–114.

25. The Dartmouth Atlas of Healthcare. http://www.dartmouthatlas.org/. Accessed April 10, 2014.

26. Birkmeyer JD, Bronner KK, Bubolz TA, et al. The Quality of Medical Care in the United States: A Report on the Medicare Program. The Center for the Evaluative Clinical Sciences Dartmouth Medical School; 1999. http://www.dartmouthatlas.org/. Accessed December 12, 2013.

27. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71–72.

28. American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers Criteria for potentially inappropriate medication use in older adults. J Am Geriatr Soc. 2012;60:616–631.

29. Silow-Carroll S, Edwards JN, Lashbrook A. Reducing hospital readmissions: lessons from top-performing hospitals. The Commonwealth Fund; 2011. http://www.commonwealthfund.org/publications/case-studies/2011/apr/reducing-hospital-readmissions. Accessed September 6, 2013.

30. Liu L, Strawderman RL, Cowen ME, Shih YC. A flexible two-part random effects model for correlated medical costs. J Health Econ. 2010;29(1):110–123.

31. Hickam DH, Weiss JW, Guise J-M, et al. Outpatient case management for adults with medical illness and complex care needs. Comparative Effectiveness Review No. 99. AHRQ Publication No.13-EHC031-EF. 2013. http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=1369. Accessed January 8, 2013.

32. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55.

33. Faries DE, Leon AC, Haro JM, Obenchain RL, eds. Analysis of Observational Health Care Data Using SAS. Cary, NC: SAS Institute Inc.; 2010.

34. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol. 2001;2:259–278.

35. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. Pharmacoepidemiol Drug Saf. 2005;14:465–476.

36. Heckman JJ, Hidehiko I, Prieti T. Matching as an econometric evaluation estimator: evidence from a job training program. Rev Econ Stud. 1997;64:605–654.

37. Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. Biometrics. 1996; 52(1):249–264.

38. Kim YJ, Soeken KL. A meta-analysis of the effect of hospital-based case management on hospital length-of-stay and readmission. Nurs Res. 2005;54:255–264.

39. Wise CG, Bahl V, Mitchell R, West BT, Carli T. Population-based medical and disease management: an evaluation of cost and quality. Dis Manag. 2006;9:45–55.

40. Committee on Quality of HealthCare in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.

41. Quinn JL, Prybylo M, Pannone P. Community care management across the continuum. Study results from a Medicare health maintenance plan. Care Manag J. 1999;1:223–231.

42. Hawkins K, Escoto KH, Ozminkowski RJ, Bhattarai GR, Migliori RJ, Yeh CS. Disparities in major joint replacement surgery among adults with Medicare supplement insurance. Popul Health Manag. 2011;14:231–238.

Address correspondence to:
Timothy S. Wells, MPH, PhD
Advanced Analytics, OptumInsight
315 E. Eisenhower Parkway, Suite 305
Ann Arbor, MI 48108

E-mail: Tim.Wells@optum.com

This article has been cited by:

1. Aaron Conway, Chris O'Donnell, Patsy Yates. 2017. The Effectiveness of the Nurse Care Coordinator Role on Patient-Reported and Health Service Outcomes: A Systematic Review. *Evaluation & the Health Professions* **12**, 016327871773461. [Crossref]

2. Wells Timothy S., Ozminkowski Ronald J., McGinn Michael P., Hawkins Kevin, Bhattarai Gandhi R., Serxner Seth A., Greame Chris. 2017. Incorporating Reporting Efforts to Manage and Improve Health and Wellness Programs. *Population Health Management* **20**:3, 181-188. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]

3. Timothy S. Wells, Ronald J. Ozminkowski, Kevin Hawkins, Gandhi R. Bhattarai, Douglas G. Armstrong. 2016. Leveraging big data in population health management. *Big Data Analytics* **1**:1. . [Crossref]

4. R. Tkatch, S. Musich, S. MacLeod, K. Alsgaard, K. Hawkins, C. S. Yeh. 2016. Population Health Management for Older Adults: Review of Interventions for Promoting Successful Aging Across the Health Continuum. *Gerontology and Geriatric Medicine* **2**:0. . [Crossref]

5. Timothy S. Wells, Gandhi R. Bhattarai, Kevin Hawkins, Yan Cheng, Joann Ruiz, Cynthia A. Barnowski, Barney Spivack, Charlotte S. Yeh. 2016. Care Coordination Challenges Among High-Needs, High-Costs Older Adults in a Medigap Plan. *Professional Case Management* **21**:6, 291-301. [Crossref]

6. Ozminkowski Ronald J., Wells Timothy S., Hawkins Kevin, Bhattarai Gandhi R., Martel Charles W., Yeh Charlotte S.. 2015. Big Data, Little Data, and Care Coordination for Medicare Beneficiaries with Medigap Coverage. *Big Data* **3**:2, 114-125. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]