Centers for Medicare & Medicaid Services

Evaluation of the Senior Risk Reduction Demonstration (SRRD) Under Medicare

Final Evaluation Report

Contract Number: RTOP CMS-08-023JF

December 20, 2013

Authors:

Betty Fout, Ph.D., IMPAQ International, LLC
Daniel Weinberg, Ph.D., IMPAQ International, LLC
Nicholas Bill, M.P.P.. IMPAQ International, LLC
Katherine Kahn, M.D., The RAND Corporation
Cary Sennett, M.D., Ph.D., IMPAQ International, LLC

Submitted to:

Pauline Karikari-Martin
CMS, 7500 Security Blvd
Mail Stop WB-06-05
Baltimore, MD 21244

Submitted by:

IMPAQ International, LLC
10420 Little Patuxent Parkway, Suite 300
Columbia, MD 21044
www.impaqint.com

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report is the final evaluation of the Center for Medicare & Medicaid Services (CMS) Senior Risk Reduction Demonstration (SRRD) under Medicare. The SRRD was intended to assess whether health risk reduction programs shown to be successful among the commercial population can be effectively adapted to the Medicare population. CMS originally contracted with five vendors with experience administering health risk reduction programs in the private sector. After the pilot testing phase, two vendors remained and carried out the demonstration. The cornerstone of these programs is a Health Risk Assessment (HRA) completed yearly by program participants. Using responses to the HRA, the vendors assessed an individual's level of risk and provided tailored health reports, customized educational materials, health counseling, self-management tools, and referrals to national and local community resources.

In 2008, CMS contracted with IMPAQ International, LLC to evaluate the SRRD. This final report presents the evaluation findings, focusing on five key areas: (1) background and overview of HRAs; (2) implementation of the demonstration by the vendors; (3) examination of the characteristics of participants; (4) the impact or effect of the demonstration on improving health and reducing risk; and (5) whether the demonstration was budget neutral. We begin this executive summary of the evaluation findings with an overview of the SRRD study design.

SRRD Design Overview

The SRRD was implemented over a three-year period, from May 1, 2009 through April 30, 2012, by two private-sector health management vendors that created new or adapted existing health promotion and disease prevention programs for Medicare fee-for-service (FFS) beneficiaries.[1] The two vendors are referred to as Vendor A and B throughout this report. Both vendors were selected by CMS based on their substantial experience successfully implementing health and wellness programs in the commercial sector.

Beneficiaries were eligible for the SRRD if they met *all* the following criteria:

- Must be 67-74 years old at the start of the demonstration,
- Must be Medicare FFS beneficiary enrolled in Parts A and B,
- Can be eligible for both Medicare and Medicaid,
- Must have Medicare as the primary payer,
- Cannot be enrolled in a Medicare Advantage Health Plan,
- Cannot be enrolled in hospice or have end-stage renal disease (ESRD),
- Cannot have resided in any institution for 100 days or more in the past 12 months, and
- Cannot have enrolled in Medicare before age 65.

[1] One vendor chose to leave the demonstration at the end of the second year.

For each vendor, 40,000 eligible Medicare beneficiaries were randomly assigned (see Figure 1), 20,000 to an Intervention Group (IG) and 20,000 to an Administrative Control Group (ACG[2]).[3] Each vendor then mailed an HRA packet to its assigned IG beneficiaries to recruit them for the study (i.e., invite them to participate). Once a vendor received a completed HRA from a beneficiary, the beneficiary was considered successfully recruited into the study and became an SRRD participant (Step 1 from Figure 1). Participants were then randomized a second time into one of three Intervention Arms for each vendor (Step 2 from Figure 1):

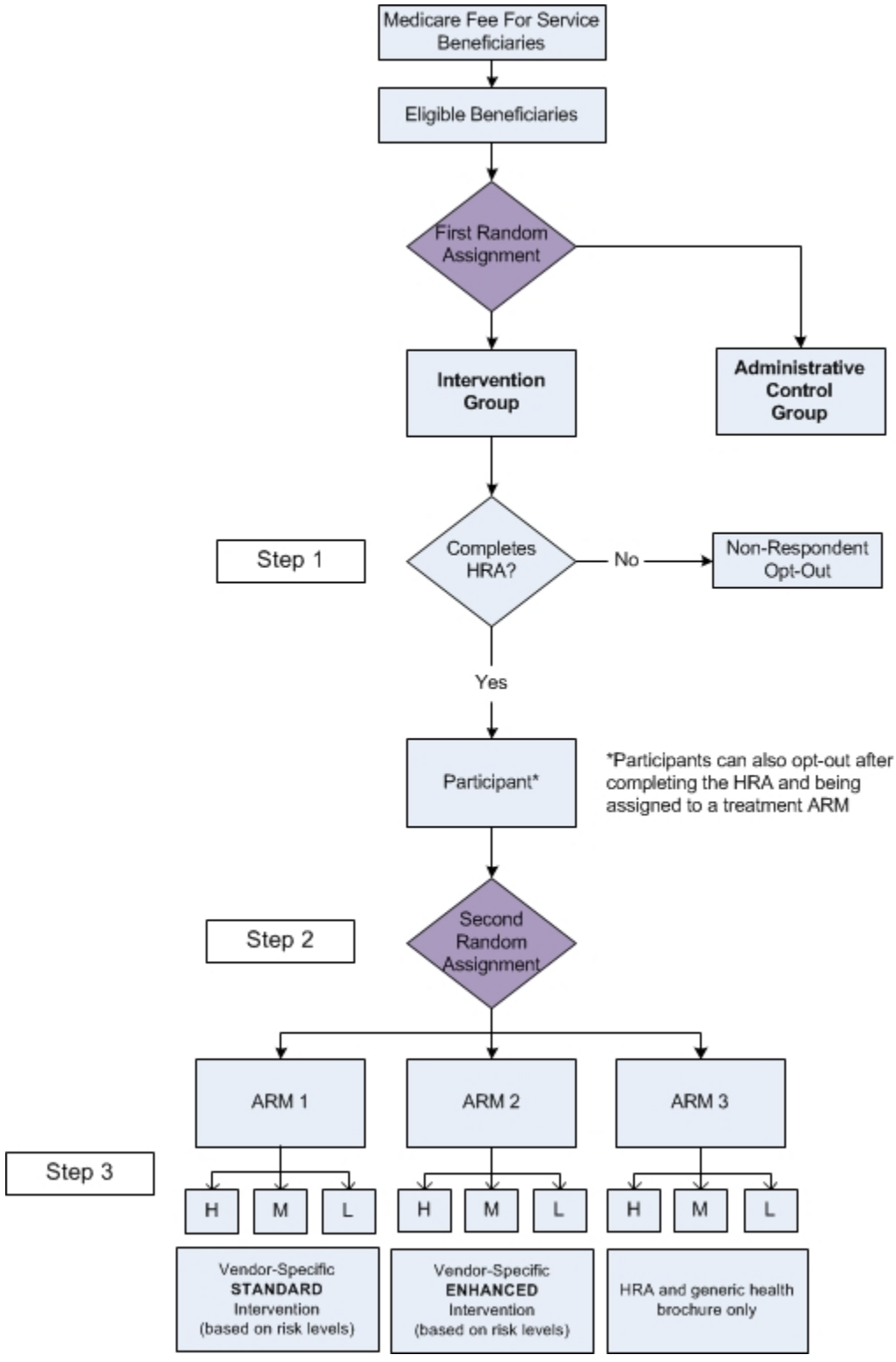
- Arm 1 – Standard Treatment (HRA + standard tailored follow-up),
- Arm 2 – Enhanced Treatment (HRA + enhanced tailored follow-up), or
- Arm 3 – HRA Only (HRA + generic health brochure).

Arm 3 involved no intervention after the HRA was completed, serving as a control group for Arms 1 and 2.

[2] The ACG was necessary to evaluate budget neutrality but did not serve as a control group for SRRD impact analyses. As discussed subsequently, Arm 3 of the IG served as the primary control in those other analyses, to isolate the effect of motivation for change without any actual intervention—as individuals randomized to Arm 3 were assumed to have similar motivation for change to those in Arms 1 and 2.

[3] Each vendor's sample was refreshed with an additional 11,800 individuals at the end of the first demonstration year.

Figure 1. SRRD Study Design



Note: H = High Risk | M = Moderate Risk | L = Low Risk

Through the HRAs, vendors collected self-reported information from Medicare beneficiaries on 17 health risk factors, specified by CMS that contribute to disease.[4] Based on the HRA responses, the vendors categorized participants into high, medium, or low risk (Step 3 of Figure 1). The HRA results were used to generate individual feedback reports, which were reviewed with beneficiaries randomized to Arms 1 and 2 (recall that Arm 3 was not an intervention group per se, but a control for Arms 1 and 2) by trained health coaches to develop plans for reducing modifiable health risk behaviors. Depending on the Arm assignment and risk category, the participants received personalized feedback and the following:

- Tailored risk reduction interventions delivered via mail, Internet, or telephone (health coaching) to program participants
- Referrals to physicians for recommended clinical preventive and screening services
- Referrals to national or local community resources and social support networks.

Vendors had discretion in the triage of participants into intervention cycles based on risk level; within each treatment Arm, participants might receive more or less intensive intervention, based on their risk. For example, a low-risk participant in the enhanced treatment Arm (Arm 2) might receive a less intensive intervention than would a high-risk participant in the standard treatment Arm (Arm 1).

The SRRD was thus designed as a randomized control trial (RCT) with two levels of randomization. At the first level, eligible Medicare FFS beneficiaries were randomly assigned to be part of the target intervention group (IG) or an administrative control group (ACG). At the second level of randomization, those in the IG who chose to participate (by returning a completed HRA) were randomly assigned into one of three Intervention Arms. These varied in program intensity, including one "control" Arm, as noted, where participants received an HRA only with no follow up interventions besides a generic health brochure. This second comparison, between treatment (Arms 1 and 2) and control (Arm 3), provides a key innovation in the evaluation of wellness and prevention programs: the "HRA only" Arm consists of individuals who chose to participate but did not receive the follow-up program, which presumably controls for the level of motivation and willingness to change—important variables that are generally difficult or impossible to account for in voluntary programs.

How was the SRRD Implemented?

Each vendor developed customized HRAs that addressed, at a minimum, the 17 CMS-specified health risk areas. The vendors used proprietary algorithms to assign participants to high-risk, moderate-risk, and low-risk categories based on the individuals' HRA responses. Interventions

[4]The following 17 health risk factors were used: physical inactivity/lack of exercise, poor nutrition, smoking/tobacco use, excessive alcohol consumption, high blood pressure, high blood glucose, high total cholesterol, being overweight/obese, inappropriate use of clinical preventive services, depression, high stress, lack of general well-being, burden of providing care giving, social isolation, lack of motor vehicle/home safety, falls (preventable accidents), and polypharmacy/medication issues.

were then tailored according to these risk categorizations. Vendors based the intensity and range of services offered on each individual's Arm assignment and risk level.

Each vendor was required to implement certain components of the demonstration as prescribed by CMS, but vendors had considerable flexibility in constructing their own risk reduction programs. The vendors developed and implemented unique approaches across several domains. These domains included participant recruitment, HRA questionnaire design and administration, risk stratification and triage, and intervention services. The IMPAQ team conducted site visits to the vendors during each year of SRRD implementation.[5] Each visit included semi-structured interviews with key vendor staff to gain an in-depth understanding of how the SRRD was designed and implemented across these different domains. Interviewees also provided their perspectives on lessons learned from the pilot and the initial implementation year. In addition, the IMPAQ team participated in interim and final evaluation meetings, during which the vendors provided additional feedback on program challenges, successes, and lessons learned. Based on the information gathered through the site visits, interviews, and meetings, we compiled the following lessons learned on SRRD program implementation:

Based on the site visits and vendor interviews, we concluded that:

- Use of a staggered/wave approach to recruitment proved a more effective strategy than sending out all recruitment materials at once.
- Careful design and branding of the SRRD recruitment materials were necessary to improve the credibility of the SRRD program to Medicare beneficiaries.
- Awareness calls and use of a CMS bridge letter in the recruitment package each year were important strategies for increasing response, participation, and re-enrollment rates.
- Uptake of the online intervention mode was limited, making the proportion of services offered online versus phone an important program consideration.
- For one vendor, switching from an "opt-in" to an "opt-out" model for program enrollment proved useful. Having discovered a lack of voluntary enrollment into health risk reduction programs after completing the HRA (opt-in model), the vendor switched to an opt-out model—in which beneficiaries were automatically enrolled into a recommended program based on data collected from the HRA. This allowed for more personalized and tailored communications with eligible beneficiaries by the health coaches.
- Not knowing beneficiaries' particular Arm assignment made recruitment difficult because vendors were unable to inform potential recruits of the exact types of services they could expect to receive.

[5] Two rounds of site visits were conducted with the vendor that did not continue into Year 3. Three rounds of site visits were conducted with the vendor that completed the demonstration.

The SRRD included both national and local components. Local samples were intended to leverage the existence of local Aging and Disability Resource Centers (ADRCs) that deliver information, referral, and assistance services.

- Through vendor and participating ADRC interviews, the ADRC staff reported that, based on conversations with local SRRD participants, the SRRD population was different from the typical ADRC client- SRRD individuals were more active and healthier, and generally expressed no interest in ADRC services.
- ADRC staff was more accustomed to handling inbound calls from a frailer population in need of services, but did adapt to the proactive approach of making outbound calls to local SRRD participants.
- The SRRD program requirement that ADRC staff initiate calls rather than responding passively to calls from beneficiaries seeking help. This proactive approach to engaging consumers proved a valuable lesson for ADRC staff, in that these ADRCs are now moving more towards advocacy-type activities (e.g., community organizing; and promotion of self-directed health care, stability, and independence).
- Despite cross-training efforts, the two vendors and the participating ADRCs faced coordination challenges due to differences in their organizational type and characteristics (such as capacity and usual consumer base).

Who Participated in the SRRD?

Although the SRRD was a RCT, the voluntary nature of the intervention required careful evaluation of a number of questions related to participation, including rates of participation across beneficiary characteristics, duration of participation, and characteristics of participating beneficiaries versus those of beneficiaries who chose not to participate. Recruiting for the SRRD was on a national scale and the program was designed to provide nationally representative results. Even so, it is important for the generalizability of the findings to the general Medicare population to understand whether those who participated were similar to or different from those who did not participate. We found that:

- Rates of participation were not significantly different for groups that differed with respect to gender, age, hospital admissions per year, or number of chronic conditions.
- Individuals belonging to minority groups, those who were dually eligible for Medicare and Medicaid, and those with total Medicare expenditures falling within the top and bottom quartiles were less likely to participate ($p<.05$).
- Individuals with cancer, osteoporosis, and arthritis were more likely to participate, while beneficiaries with the chronic conditions of Alzheimer's disease or stroke/Transient Ischemic Attack (TIA) were less likely to participate ($p<.05$). The latter result is consistent with the SRRD requirement that beneficiaries have sufficient mental capacity to participate.

- In general, the characteristics of beneficiaries who participated in the demonstration for at least one year were statistically indistinguishable from those who participated for more than one year. The only exception was for race/ethnicity: multi-year participants were more likely to be non-Hispanic whites ($p<.05$).
- Participants assigned to the "HRA-only" Arm participated longer on average than did participants assigned to the more intensive treatment Arms ($p<.01$). This finding could be because either (1) beneficiaries found participation in the more intensive intervention burdensome or (2) the benefits of the intervention accrued quickly (so that wellness goals were achieved quickly) and participants considered continued enrollment to have small additional benefit.

Although SRRD eligibles were already a healthier group of Medicare beneficiaries (due to eligibility requirements that, for example, excluded ESRD and institutionalized beneficiaries), participants were still among the healthier of the eligible beneficiaries. Also, white non-Hispanics participated at higher rates than did other races and ethnicities, suggesting that the impact of the demonstration may not be generalizable to the broader population. We did find that individuals who chose to re-enroll for multiple years were generally similar to those who participated for only one year with respect to all individual characteristics except race/ethnicity- multi-year participants were more likely to be non-Hispanic whites.

What Were the Impacts of the SRRD on Claims-based Health Outcomes?

Medicare data provide a rich source of information on the health care use of beneficiaries in this study. We linked eligible beneficiaries to their Medicare claims history, allowing us to access beneficiaries' Part A and Part B Medicare claims information from a baseline period through the end of the demonstration. For each vendor, we compared outcomes constructed with the health claims data for beneficiaries randomized into treatment and control groups. The SRRD design allows for comparisons to examine numerous aspects of the program, including variation in program impact by year, variations in impact within the National and Local samples, and variation in impact related to intensity of intervention.

We examined the impact of the SRRD on Medicare expenditure and use variables as well as certain preventive screenings. We found:

- Statistically significant decreases in total Medicare and hospital inpatient expenditures in Year 1 of the demonstration for Vendor A. These impacts were not sustained in Year 2.
- Evidence of lower spending, lower use, and higher preventive screening rates among Vendor B's local sample in Year 2 of the demonstration. These impacts were not sustained in Year 3.

Thus, the SRRD did not have lasting impact across the claims-based health outcomes we examined, even with the use of several alternative unbiased estimators. However, given the nature of the risk reduction programs and the potential pathways through which the SRRD would likely operate, decreases in Medicare costs during the relatively short period of the

demonstration may not be likely. We therefore examined alternative pathways for SRRD impact, such as the effects of the programs on beneficiaries' risk behaviors.

What Was the Impact of the SRRD on Participants' Risk?

Using the HRA data, we examined the impact of the SRRD on risky behaviors, providing valuable information not captured by claims data for the study period. The HRAs also potentially provide insight into the pathways through which the SRRD operates. For example, SRRD-induced improvements in behavior (e.g., improved preventive care or reduced stress) might take years to translate into the changes in costs and use that can be detected in health claims data. Comparing individuals' HRA data pre- and post-intervention allowed us to investigate whether the SRRD was associated with changes in health risk behaviors likely linked to improved health outcomes that cannot be observed in the time frame available for the Medicare claims data.

The purpose of the risk reduction programs implemented by the SRRD vendors was to improve health through the identification of risky behaviors (via the HRA) and to tailor an appropriate follow-up program that would motivate and enable participants to alter their behavior to reduce risk and ultimately improve their health. We examined the impact of the SRRD on participants' risk scores constructed by the vendors using participants' HRA responses, as well as the impact of the SRRD on self-reported health status measures from the HRA. Our methodology allowed us to account for selection factors that might bias our analysis, since we could observe HRA responses over time only for those beneficiaries choosing to participate in the program for more than one year. Results showed that the SRRD was, indeed, successful in reducing risk in certain areas:

- Beneficiaries participating in Vendor A's risk reduction programs showed statistically significant improvements in the risk areas of stress and general well-being. In addition, we detected an improvement in overall risk for Vendor A's participants.
- Beneficiaries participating in Vendor B's risk reduction programs showed statistically significant improvements in the risk areas of back care, nutrition, physical activity, stress, and general well-being as well as for overall risk.

Moreover, participants in both vendors' programs showed improvements in self-rated health status. For Vendor B, we were also able to assess whether longer participation in the risk reduction programs resulted in greater impacts. Beneficiaries who participated for all three years of the demonstration had the largest decreases in the probability of being categorized as high risk in the same areas listed above, as well as greater improvements in self-rated health status.

Was the SRRD Budget Neutral?

To implement the SRRD, CMS paid vendors a fee for each beneficiary recruited into the vendors' programs. CMS deferred paying a portion (10 percent) of that fee pending the results of a budget neutrality analysis. We used three different methodologies to assess whether the programs, including fees paid to the vendors, were budget neutral or added expenditures to the Medicare program. The methodologies compared overall Medicare costs for beneficiaries in the IG to Medicare costs for those in the ACG over the duration of the demonstration, and yielded the following findings:[6]

-Vendor A results were budget neutral using all three methods.

-Vendor B results were budget neutral using one method only.

Caution is required, however, in drawing conclusions on the financial viability of the SRRD based on these results. The final budget neutrality calculation is an accounting activity that uses a pre-established methodology to determine whether deferred funds should be retained or distributed to the vendors. The budget neutrality calculation does not consider the statistical significance of any apparent differences in spending. In fact, the calculated net costs for both vendors were not statistically significant for any of the three methodologies—that is, Medicare spending for beneficiaries in the IG was not statistically different from those in the ACG for either vendor.

Lessons Learned and Recommendations

Until recently, CMS offered beneficiaries few programs designed to reduce risks of preventable illnesses or to promote healthy lifestyles. However, if Medicare invests in the health and well-being of its beneficiaries by offering comprehensive and effective health promotion and disease prevention programs, seniors may reduce their modifiable health risks, have fewer health problems, improve their quality of life, forestall disability, consume fewer health care resources, and be more productive. The 2010 Patient Protection and Affordable Care Act included: 1) provisions intended to prevent the onset of preventable chronic conditions and improve the overall health of Americans and 2) support for the delivery of evidence-based prevention and wellness services. In January 2011, Medicare began covering, without cost to beneficiaries, an annual wellness visit focused on prevention and screening; as of January 2012, use of an HRA became a required part of the annual wellness visit.

The SRRD provides a timely opportunity to examine the effects of implementing an HRA-based health risk reduction program for the Medicare population. Under SRRD, private sector vendors administered low-cost HRA-based health risk reduction programs to eligible Medicare FFS beneficiaries. The programs were intended to maintain the health and reduce health risk of

[6] Recall, Medicare beneficiaries were randomized to the ACG or to the IG (which included participants of Arms 1, 2, and 3 as well as those who were invited to, but did not, participate); the ACG was designed so that we could evaluate the budgetary implications of the SRRD.

eligible beneficiaries. The three-year demonstration was rigorously designed as an RCT and intended to provide nationally representative results.

The vendors reported facing challenges to engaging the SRRD population. First, awareness-building that occurs at the worksite for employer-based wellness programs was not possible for the Medicare population. Second, not knowing during initial recruitment which actual treatment group the participant was assigned to was a barrier to recruitment, because vendor staff were not able to explain to beneficiaries what specific services they would receive and, thus, what benefits program participation would bring.

However, vendors were ultimately able to meet recruiting targets. They found practices such as yearly awareness calls and yearly letters with careful CMS branding were useful in providing reminders and ensuring trust. These practices may also prove useful in reminding Medicare beneficiaries of the availability of the new Medicare annual wellness visit with no out-of-pocket cost. Only about 9 percent of Medicare beneficiaries have so far made use of this annual visit.[7] One suggested reason for the low take-up rates is the difficulty in making beneficiaries aware of and understand this benefit.

Individuals who were successfully recruited into the SRRD, as noted, were more likely to be healthier and to be a non-Hispanic white, and less likely to be dually eligible. Since vendors were not provided information regarding their target populations, they were not able to tailor recruitment materials and practices to the individual beneficiary. Thus, it may be important to consider more focused efforts to reach the less healthy, lower income, and minority populations, for which significant potential for mitigating health risks likely exists.

The design of the HRA may also play an important role in recruiting individuals to complete the form and participate in the program. One vendor designed the HRA to be understandable at a 6th grade reading level, developed questions specific to the older population, conducted focus group testing, and was able to recruit non-Hispanic whites and dual eligible beneficiaries at higher rates. Another vendor, who developed a 7th grade reading-level HRA questionnaire with questions adapted from its existing HRAs used for the working-age population, had lower rates of non-Hispanic white and dual eligible participation.

For those who did participate, the SRRD was successful in decreasing the probability of being categorized as high risk for certain risk areas (such as stress, back care, smoking/tobacco use, nutrition, physical activity, and overall well-being). Risk areas where we found no impact were motor safety, weight, polypharmacy, and falls. The risk reductions and overall self-reported health improvements did not translate into lower total Medicare expenditures or Medicare expenditures for certain types of care during the study period. However, the study period covered only the time during which the demonstration was operating (three years); it is possible that risk reductions would appear in the cost data over a longer period (both a longer study period and a longer follow-up period).

[7] http://www.marketwatch.com/story/medicare-wellness-visits-worth-a-second-look-2013-02-26

Finally, the vendors recommended "bridging" between the physician's office and the program. Under the SRRD, vendors' health coaches operated outside the clinical setting, similar to workplace wellness program models, which may be a less costly than those that are integrated with the physician's office. However, vendors felt that connecting the program to the health care provider would have improved patient engagement. This potential "hybrid" approach—using health coaches in coordination with the physician's office—is a potential opportunity for future investigation, particularly in conjunction with the Medicare annual wellness visit.

1. BACKGROUND OF SRRD

Behavior lifestyle choices regarding diet, exercise, and the use of alcohol and tobacco are associated with several chronic diseases as well as leading causes of mortality in the United States.[8] In recent years, research has suggested that well-designed, HRA-based risk reduction programs implemented in the workplace can encourage and help sustain the types of behavioral changes that may prevent future health problems—thus lowering the risk profile of the individual participant, increasing overall health, and potentially lowering health costs.[9]

In the past, CMS offered beneficiaries limited programs designed to reduce risks of preventable illnesses and/or to promote healthy lifestyles. In recent years, however, interest has risen in programs that show promise for increasing quality and years of healthy life as set forth in the Healthy People 2020 vision by the Department of Health and Human Services (DHHS). Moreover, research studies—including the *Evidence Report and Evidence-Based Recommendations: Health Risk Appraisals and Medicare*, commissioned by CMS and prepared by RAND—have concluded that effective risk reduction programs can improve general health status outcomes and may produce a positive return on investment, and have recommended testing their use among the Medicare population.[10] The SRRD provides an opportunity to learn how best to adapt effective HRA-based risk reduction programs to that population.

1.1 Overview of Health Risk Assessments

HRAs have been used for several decades, most widely in health education and health promotion programs in the workplace.[11] Originally, HRAs were used to collect health risk data from individuals to produce personalized epidemiological-based projections of mortality risk.[12] HRAs have since evolved to become interactive, web-based tools that provide individualized feedback and educational messages designed to motivate behavior change and risk reduction. HRAs may include the following general elements:[13]

- Assessment of personal health habits and risk factors supplemented with biometric measurements of physiologic health
- Quantitative estimation or qualitative assessment of future risk of death or adverse health outcomes

[8] http://www.cdc.gov/chronicdisease/overview/index.htm

[9] Baicker, Katherine, Cutler, David, Song, Zirui, "Workplace Wellness Programs Can Generate Savings," *Health Affairs*, 29(2), 2010.

[10] *RAND. (2001). Evidence Report and Evidence-Based Recommendations: Health Risk Appraisals and Medicare.* Baltimore, MD: U.S. Dept. of Health and Human Services, Health Care Financing Administration.

[11] Soler, R.E. et al. (2010). A systematic review of selected interventions for worksite health promotion: the assessment of health risks with feedback. *Journal of Preventive Medicine*, 38(2S), S237–S262.

[12] Schoenback, V.J., Wagner, E.H., and Beery, W.L. (1987). Health risk appraisal: review of evidence for effectiveness. *Health Services Research,* 22(4), 553–580.

[13] Partnership for Prevention and Thomson Reuters (2011). General Proceedings from a Public Forum, Expert Input, and the Research Literature for the Design of Patient-Centered Health Assessments, Final Report, March 15, 2011.

- A mechanism for providing feedback in the form of educational messages or counseling about ways to change behavior and health habits to potentially alter risk of disease or premature death.

HRAs can be important tools for raising awareness of health issues to encourage behavioral change, as well as for triaging individuals into risk-appropriate interventions and tracking changes over time. However, they have limitations, which include inaccuracy of the information; recall bias; respondents' lack of understanding of health risk questions; and the need to tailor HRAs to specific literacy, cultural, and age groups. Furthermore, findings show that HRAs alone are not effective in inducing long-term behavior change. Further support is necessary. Thus, HRAs should be considered a first step towards a comprehensive framework of behavioral change and risk reduction.[14]

In January 2011, as specified in Section 4103 of the 2010 Affordable Care Act (ACA), Medicare began covering, without cost to beneficiaries, an annual wellness visit focused on prevention and screening.[15] As of January 2012, use of an HRA became a required part of the annual wellness visit.[16] As part of the 2010 law, the Secretary of DHHS was authorized to establish publicly available guidelines for HRAs, after consultation with relevant groups and entities.[17] Five specifications were listed for development of those guidelines. They should (1) identify chronic diseases, injury risks, modifiable risk factors, and urgent health needs of an individual; (2) be furnished through an interactive telephonic or web-based program; or (3) be offered during the encounter with a health care professional or through community-based prevention programs; or (4) be provided through any other means the Secretary determines appropriate to maximize accessibility and ease of use by beneficiaries.[18] The Centers for Disease Control and Prevention (CDC) has provided guidance regarding, HRAs and their modes of provision for Medicare beneficiaries.[19] General guidance includes the following:

- All questions in the HRA must be actionable.
- Feedback received by the patient from the provider regarding HRA results should be the result of shared decision making.
- The HRA should be written at a 5th or 6th grade literacy level and in plain language.

[14] Partnership for Prevention and Thomson Reuters (2011).

[15] Koh, H.H. and Sebelius, K.G. (2010). Promoting prevention through the Affordable Care Act. *New England Journal of Medicine*, 363(14), 1296–99. The use of a health risk assessment became a required part of the Annual Wellness Visit in January 2012.

[16] https://cmsadmin30.convio.net/preview!www.ncoa.org/assets/files/pdf/center-for-benefits/medicare-health-risk-assessment.pdf?authToken=2de829aeb1cb787715bc1a04efef4649faf260b7

[17] Partnership for Prevention and Thomson Reuters (2011).

[18] Goetzel, Ron Z., et al. A Framework for Patient-Centered Health Risk Assessments: Providing Health Promotion and Disease Prevention Services to Medicare Beneficiaries. Available at: http://www.cdc.gov/policy/opth/hra/FrameworkForHRA.pdf

[19] Centers for Disease Control and Prevention (2011). Interim Guidance for Health Risk Assessments and Their Modes of Provision for Medicare Beneficiaries.
Available at http://prevent.org/data/files/news/healthriskassessmentscdcfinal.pdf.

- The HRA should be linguistically, age, gender, and culturally appropriate for the patient.
- The HRA should be received no less than every two years to ensure compliance with current science related to health promotion and disease prevention and to take advantage of advances in technology.

In addition, the CDC identified six areas that the HRA and HRA delivery should address: (1) content and design, (2) mode of administration, (3) primary care office capacity, (4) consumer/patient perspectives, (5) data, and (6) evaluation and quality assurance. The CDC has also released an "evidence-informed" framework for providers, policymakers, health plans, payers, researchers, and vendors for implementing HRA-based preventive health programs for Medicare beneficiaries. Specific CDC recommendations for an improved "HRA *Plus*" process include, in addition to administering a HRA and providing a feedback report:[20]

- Multiple HRAs, with feedback provided over time
- Ongoing health education programs through pamphlets, books, videos, or interaction computer programs
- In-person or telephone counseling or coaching to support behavior change and risk reduction
- Referral to community resources
- Referral to local or national health promotion vendors and services (e.g., smoking cessation phone lines and wellness coaches).

The CDC has also provided a sample HRA,[21] though this HRA does not contain all the elements specified by CMS' final rule definition of HRAs.[22]

In this current landscape, the SRRD evaluation is of particular interest. The SRRD, as noted, tests whether HRA-based health promotion and health management programs developed and tested in the private sector can be tailored to and work well with Medicare beneficiaries to improve their health and reduce avoidable health services use. Many aspects of CDC's guidance and recommendations have been incorporated by the two vendors in this demonstration.

[20] Goetzel, RZ; Staley, P; Ogden, L; Stange, P; Fox, J; Spangler, J; Tabrizi, M; Beckowski, M; Kowlessar, N; Glasgow, RE, Taylor, MV. A framework for patient-centered health risk assessments – providing health promotion and disease prevention services to Medicare beneficiaries. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, 2011. Available at: http://www.cdc.gov/policy/opth/hra/.

[21] Ibid

[22] Medicare Program. Payment Policies Under the Physician Fee Schedule, Five-Year Review of Work Relative Value Units, Clinical Laboratory Fee Schedule: Signature on Requisition, and Other Revisions to Part B for CY 2012. *Fed Regist*. 2011;76(228):73306. http://www.gpo.gov/fdsys/pkg/FR-2011-11-28/pdf/2011-28597.pdf. See also: http://www.acponline.org/running_practice/payment_coding/medicare/2012_mpfs_rule.pdf

1.2 Evaluation of the SRRD

Given the focus placed on preventive care and promoting good health through wellness programs under the ACA, HRA-based risk reduction programs are continuing to gain popularity. With the expansion of these programs, especially in the workplace, several studies have brought to light the difficulty in evaluating their impacts and questioned the reliability of studies finding large positive benefits.[23] Since individuals select to participate into these programs, obtaining a comparison group of individuals who are similar to participants but did not participate (a crucial step for conducting a rigorous evaluation of the program) can be an insurmountable challenge.

The SRRD was a uniquely designed randomized control trial (RCT) with two levels of randomization. At the first level, eligible Medicare fee-for-service (FFS) beneficiaries were randomly assigned to be part of the target intervention group (IG) or an administrative control group (ACG). At the second level of randomization, those in the IG who chose to participate (by returning a completed HRA) were randomly assigned into one of three treatment arms. These varied in program intensity and included one "control" Arm, in which participants received an HRA only with no follow-up interventions. This second comparison, between treatment and control Arms, provides a key innovation in the evaluation of wellness and prevention programs. The "HRA only" Arm consists of individuals who chose to participate but did not receive the follow-up program, and presumably controls for the important variables of motivation level and willingness to change, which are normally difficult or impossible to account for in voluntary programs.

[23] Mattke, Soeren, Schnyer, Christopher, Van Busum, Kristin, "A Review of the U.S. Workplace Wellness Market," RAND Occasional Paper, July 2012. Also see http://healthaffairs.org/blog/2013/01/16/is-it-time-to-re-examine-workplace-wellness-get-well-quick-schemes/

2. SRRD PROGRAM IMPLEMENTATION EVALUATION

2.1 Introduction

In this chapter, we provide an overview of the SRRD design and examine its implementation by the two health management vendors selected by CMS to carry out the demonstration. The information is based on reviewing the SRRD design provided by CMS and its implementation contractor, as well as site visits and interviews conducted with the vendors.[24]

Summary of Findings

- The SRRD was implemented by two health management vendors that developed tailored risk reduction programs based on HRAs for Medicare beneficiaries nationwide and in selected local areas.
- Each vendor generally met recruitment goals using a range of recruitment strategies.
- Use of a staggered/wave approach to recruitment proved to be a more effective strategy than sending out all recruitment materials at once.
- The design and branding of the SRRD recruitment materials were important for improving credibility of the SRRD program to Medicare beneficiaries.
- There was limited uptake of risk reduction services offered through web-based modes.
- An opt-out strategy for enrolling participants into the program proved to be more successful than an opt-in strategy.
- In the local SRRD component, vendors partnered with local Aging and Disability Resource Centers (ADRCs) to provide information, referral, and assistance services (IR&A):
  - The SRRD population was more active and healthier than typical ADRC customers.
  - The SRRD population was generally not interested in ADRC services.
  - ADRC staff appreciated skills learned from the proactive approach to engaging consumers, since they are moving more towards advocacy-type activities.

2.2 SRRD Design

The SRRD was implemented over a three-year period, from May 1, 2009 through April 30, 2012, by two private-sector health management vendors that created new or adapted existing health promotion and disease prevention programs for Medicare beneficiaries. Preceding the three-

[24] The implementation contractor was Thomson Reuter (now called Truven).

year demonstration, the vendors participated in two pilot implementation phases.[25] CMS originally selected five vendors to implement the SRRD, but after participating in the pilot three of the vendors decided not to continue. After Year 2 of the demonstration, one of the two remaining vendors chose to withdraw, leaving only one vendor to implement the final Year 3 of the demonstration.

The two vendors are referred to as Vendor A and B throughout this report. Both were selected by CMS and have substantial experience successfully implementing health and wellness programs in the commercial sector. Vendor A has an extensive history of partnering with health care and community organizations to implement patient-centered programs that focus on prevention, disease management, and care coordination to improve patient health and health care delivery. To implement the demonstration, Vendor A partnered with a leading provider of employee health improvement services to Fortune 500 companies, the health care industry, and individual consumers for over 30 years. Vendor B has been in business for over 30 years and was one of the first health and wellness corporations. Vendor B has over 500 employees and markets its services to Fortune 1000 companies. Most of Vendor B's clients are employers with 7,500 or more employees in industries such as utilities, manufacturing, education, commerce, banking, technology, among others.

The SRRD consisted of the following core program components:

- A vendor-administered initial HRA followed by a tailored feedback report
- Randomization of participants into a set of risk reduction modules (treatment Arms), with varying intensity of intervention services
- Participant risk stratification based on HRA results and proprietary algorithms developed by each vendor
- Provision of tailored risk reduction interventions delivered via mail, Internet, or telephone (health coaching) to program participants
- Referrals to physicians for recommended clinical preventive and screening services
- Referrals to national or local community resources, social support networks, and volunteer opportunities.

The SRRD was designed to have two parallel demonstrations: (1) a core National demonstration, which used a random selection of beneficiaries from across the U.S. and (2) a Local demonstration, which used a random selection of beneficiaries from communities with exemplary information, referral, and assistance (IR&A) providers.[26] In addition to tailored health risk reduction interventions, beneficiaries in the National sample received referrals to

[25] The enrollment period for Pilot 1 began April 1, 2008 and lasted through June 23, 2008. The services to these beneficiaries extended for 12 months, through March 31, 2009. The enrollment period for Pilot 2 began September 1, 2008 and lasted through December 31, 2008. The beneficiaries from this pilot were rolled into the demonstration.

[26] Identified by the National Council on the Aging (NCOA) and the Administration on Aging (AoA).

national resources and organizations. SRRD participants in the Local samples received tailored referrals to resources and organizations in their local area, in addition to national ones. The purpose of including the local demonstration component was to assess how Medicare beneficiaries who received referrals to local services fared in comparison to beneficiaries who received only referrals to nationally available resources.

Each health management vendor was assigned two Local samples and was responsible for coordinating the implementation of the local SRRD with the ADRC in those communities. Prior to the official start of the demonstration on May 1, 2009, ADRC and vendor management staff assigned to the SRRD met regularly via on-site visits or conference calls to conduct planning activities, discuss the goals and objectives of the demonstration, and educate each other on the roles and responsibilities of each. The basic role of the ADRCs in the demonstration was to conduct outreach to local SRRD participants and provide tailored referrals to community resources, social support networks, and volunteer opportunities. Once a beneficiary in the Local sample became an SRRD participant by completing and returning an HRA, the vendor was responsible for notifying the ADRC and providing the participant's contact information. A new list of local SRRD participants was provided to each ADRC on a weekly basis. Upon receiving the list, the ADRC staff conducted outreach to each local SRRD participant to offer their services and provide localized referrals.

Figure 1 illustrates the first step in the general SRRD study design. Medicare FFS beneficiaries were selected for the SRRD based on the following eligibility criteria:

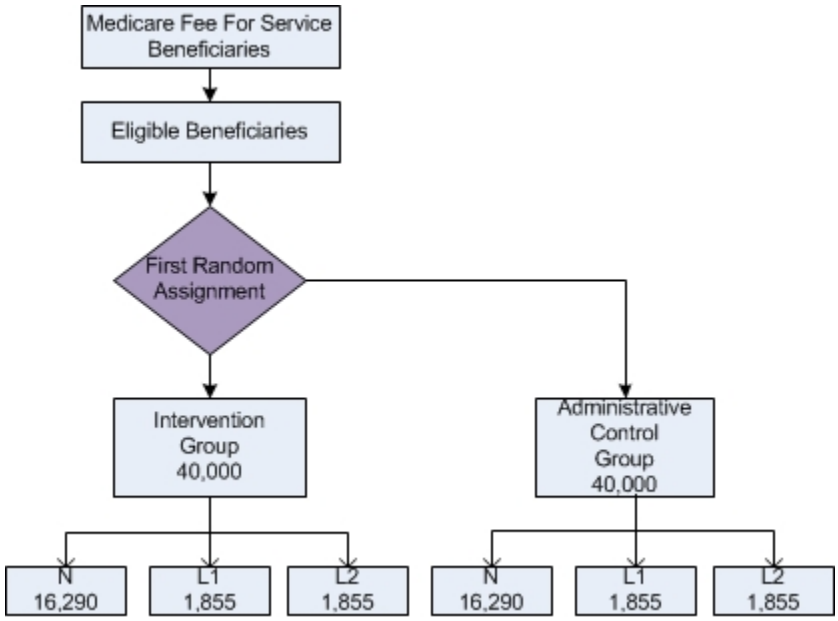
- Must be between 67 and 74 years old at the start of the demonstration
- Must be a Medicare FFS beneficiary enrolled in Parts A and B
- Can be eligible for both Medicare and Medicaid
- Must have Medicare as the primary payer
- Cannot be enrolled in a Medicare Advantage Health Plan
- Cannot be enrolled in hospice or have end-stage renal disease (ESRD)
- Cannot have resided in any institution for 100 days or more in the past 12 months
- Cannot have enrolled in Medicare before age 65.

For each vendor, eligible beneficiaries were then randomized into the IG or ACG comprising three samples: two Local samples (L1 and L2) and a National sample (N).[27] Each SRRD health management vendor was randomly assigned 20,000 IG beneficiaries (16,290 + 1,855 + 1,855 =

[27] For the National component, zip codes across the country were random assigned to each vendor and beneficiaries were randomly sampled from the assigned zip codes. For the Local samples, beneficiaries were randomly sampled from the zip codes corresponding to the two ADRCs each vendor was assigned. Zip codes that were part of the Local samples were excluded from the National component. CMS conducted eligibility checks on all sampled beneficiaries. Ineligible beneficiaries were dropped and replaced with a randomly selected beneficiary from the same zip code.

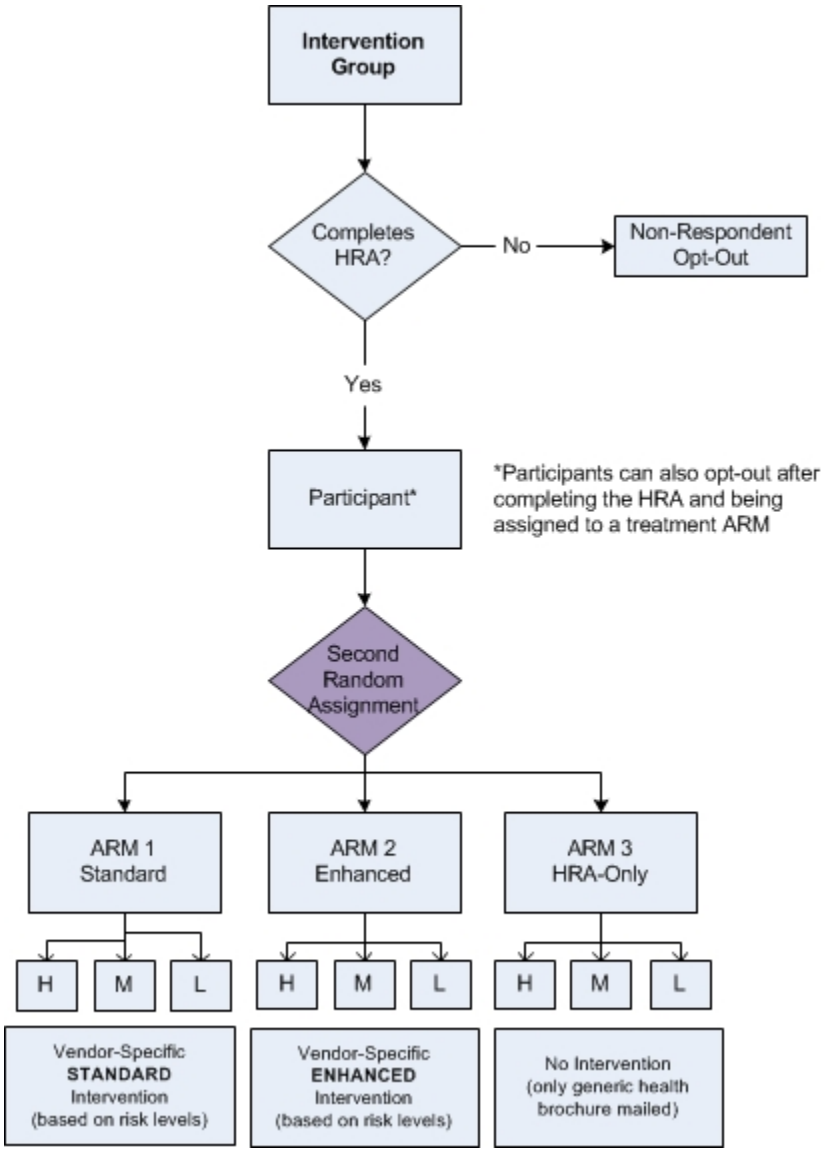
20,000) split among the respective sample types (L1, L2, and N). Each vendor was also randomly assigned 20,000 ACG beneficiaries (split among the sample types). Thus, each vendor was randomly assigned 40,000 beneficiaries (20,000 IG + 20,000 ACG).

Figure 1: First Step in the SRRD Study Design



Next, the vendors received the contact information of the individuals randomized into the National and Local samples of the IG for recruitment into the study. Members of the ACG were not contacted (see Figure 2). The vendors sent recruitment materials, including an HRA packet, to their assigned IG beneficiaries to recruit them for the study. Once a vendor received a completed HRA, the beneficiary was considered successfully recruited into the study and was termed an SRRD participant. Once recruited, each SRRD participant was then randomly assigned by the implementation contractor to one of three Intervention Arms.

Figure 2: Second Step in the SRRD Design



Depending on the assignment, the participant received personalized feedback and/or a health coaching intervention. The demonstration design required that each vendor administer the following three levels of intervention:

- *Arm 1* – Standard Treatment (HRA + standard tailored follow-up),
- *Arm 2* – Enhanced Treatment (HRA + enhanced tailored follow-up), or
- *Arm 3* – HRA Only (HRA + generic health advice).

Each of these levels of intervention is described below.

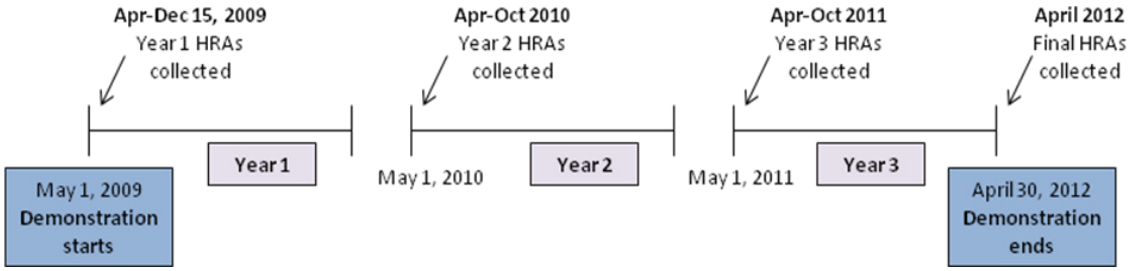
- *Arm 1 – Standard Treatment (HRA + standard tailored follow-up)*: Arm 1 participants received the standard intervention. Under the standard intervention, vendors provided the following services: an individualized feedback and follow-up report; tailored, behavioral risk-specific intervention modules delivered through the mail, via the Internet (if the participant preferred), or (optionally) through proactive telephone counseling and health coaching; high-risk programming;[28] a help line that participants could call with questions and concerns; and referrals to national or local risk reduction resources.
- *Arm 2 – Enhanced Treatment (HRA + enhanced tailored follow-up)*: Arm 2 participants received an enhanced intervention. Arm 2 provided all the components of Arm 1 but also offered more intensive interventions, including required proactive telephone counseling for a subgroup of participants. The individual vendor determined which participants were most suitable for telephone counseling. In addition, vendors could offer additional tailored behavior change modules, more frequent and/or more intensive interactions with beneficiaries, and greater access to health educators to support risk reduction efforts.
- *Arm 3 – HRA Only (HRA + generic health advice)*: Arm 3 was designed as the placebo intervention. Participants assigned to Arm 3 received an HRA but no intervention aside from receiving a standardized, non-tailored letter describing the advantages of a healthy lifestyle. No additional follow-up interventions other than a generic health brochure were provided. The vendors assigned risk scores to the participants based on the HRAs, but the participants did not receive their HRA results.

Vendors had significant discretion in the triage of participants into intervention cycles based on risk level. Within Arms 1 and 2, participants might receive several intensities of intervention. For example, a low-risk participant in the enhanced treatment Arm 2 might receive a less intensive intervention than would a high-risk participant in the standard treatment Arm 1.

Recruited beneficiaries (participants) completed an HRA between April 1, 2009 and December 15, 2009 (the beginning of Year 1). Additional HRAs were collected at the beginning of Year 2 and Year 3, and a final Exit HRA was collected at the end of the demonstration. Thus, four rounds of HRAs were collected during the demonstration, with the Year 1 HRA considered the HRA taken at "baseline" (see Figure 3).

[28] Vendors could offer additional, more frequent, or more intensive interventions to participants categorized as high risk.

Figure 3: Timing of HRA Collection [a]



[a] One vendor chose not to continue with the demonstration after Year 2. Therefore, the Year 3 and final HRAs were collected only for one vendor's participants.

The HRAs used in the SRRD collected self-reported information from Medicare beneficiaries on 17 health risk factors specified by CMS that contribute to disease.[29] The HRA results were used to generate individual feedback reports, which were then reviewed with the beneficiaries by trained health coaches to develop plans for reducing modifiable health risk behaviors (excluding Arm 3).

2.3 Vendors' Process and Findings

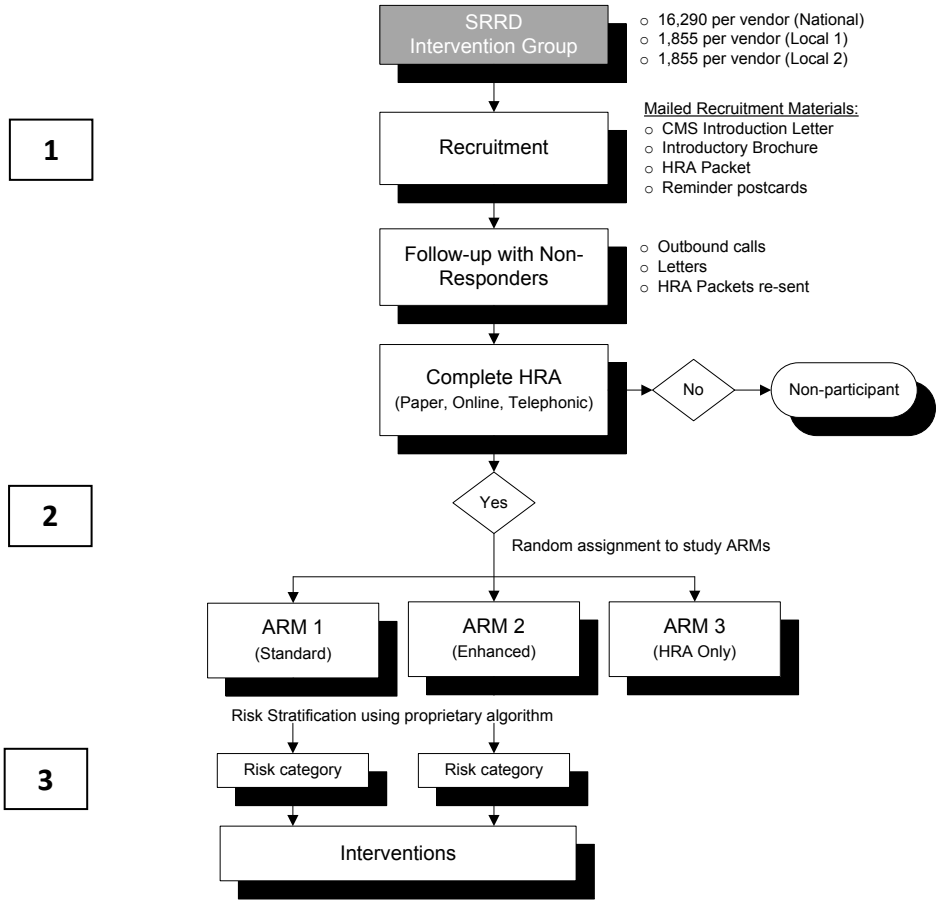
In this section, we provide an in-depth description of how the different demonstration components were implemented by the two health management vendors.

Although each vendor was required to implement certain components of the demonstration as prescribed in the demonstration design and its terms and conditions, considerable flexibility and innovation were allowed to vendors in constructing their own risk reduction programs. Each vendor developed and implemented its own approaches to participant recruitment, HRA questionnaire design and administration, risk stratification and triage, and intervention services.

Figure 4 provides a high-level flowchart outlining the key components of the process applied to members of the IG assigned to each vendor (members of the ACG were not contacted by or available to the vendors). These key components were: (1) the recruitment process, including follow-up with non-responders, (2) HRA completion and random assignment to one of the three SRRD study arms, and (3) risk stratification and the delivery of health risk reduction interventions by vendors. In this section, we describe the specific approaches used by the vendors to address each component.

[29] The complete 17-item list of CMS-specified health risk factors is: physical inactivity/lack of exercise, poor nutrition, smoking/tobacco use, excessive alcohol consumption, high blood pressure, high blood glucose, high total cholesterol, being overweight/obese, inappropriate use of clinical preventive services, depression, high stress, lack of general well-being, burden of providing care giving, social isolation, lack of motor vehicle/home safety, falls (preventable accidents), and polypharmacy/medication issues.

Figure 4: Vendor Process Flowchart



2.3.1 Recruitment and Retention

For both vendors, the first stage of SRRD implementation required inviting Medicare beneficiaries in the SRRD National and Local IG sample pools to participate in the study. The recruitment process began with the mailing of materials, including a CMS Introduction Letter, Introductory Brochure, and HRA packet. Following these mailings, each vendor made efforts to increase the response rate by attempting to contact beneficiaries who did not complete and return an HRA.

Vendor A. In Year 1 of the demonstration, Vendor A decided to use a wave approach for recruitment. The recruitment materials were mailed in six different waves and consisted of an initial CMS packet with a cover letter and endorsement letter from C. Everett Koop, M.D., a "Get Ready" postcard that was oversized to grab attention, and an introduction packet with an SRRD program cover letter, HRA questionnaire with return envelope, privacy statement, and frequently asked questions document. Reminder postcards were sent as a fourth mailing to non-respondents. Each mailing wave was followed by up to five awareness calls conducted by

customer service staff to non-respondents. Vendor staff conducted an average of three calls to each non-respondent in Year 1.

In Year 2 of the program, Vendor A improved its recruitment methods to help it stay on target with respect to enrollment goals. The vendor used the same wave approach for recruitment, but decided to reach out first to the participants from Year 1 and then to new beneficiaries, a strategy that may have contributed to achieving a higher re-enrollment rate. Because the vendor had some initial concerns about participant retention, a CMS "bridge letter" was also developed and sent to the Year 1 beneficiaries approximately eight weeks prior to the recruitment period, informing them that the Year 2 HRA was forthcoming. For Year 2, the vendor reported that fewer telephonic outreach (awareness calls) were required to reach enrollment targets.

Vendor A decided not to continue with Year 3 of the demonstration and developed a letter, which was reviewed and approved by CMS, informing all SRRD participants that the demonstration would be ending and thanking them for their participation.

Vendor B. In Year 1 of the demonstration, Vendor B sent its recruitment materials to its entire SRRD IG sample (20,000) in a single wave. The CMS packet was mailed first and was followed by outbound awareness calls, half of which were automated using Interactive Voice Response (IVR) technology and the other half made in person using outsourced staff. The introductory brochures and HRA packets were mailed out after the first round of awareness calls were completed. After the HRA packets were sent, reminder postcards were sent and a second round of awareness calls was conducted. Vendor staff conducted up to two awareness calls to each beneficiary. The maximum number of attempts for the National sample and for one of the Local samples was six. For Vendor B's other Local sample, up to eight calls were made to each beneficiary to meet enrollment targets because that particular population posed unique recruiting challenges.

In Year 2, Vendor B decided to change its recruitment approach and sent out the materials and conducted awareness calls in four waves of 5,000 beneficiaries each. This change enabled the vendor to handle the volume more efficiently, improved the follow-up time with beneficiaries, improved the response rate, and made the overall recruitment process more fluid. The vendor also decided to conduct all awareness calls using the IVR system (rather than outsourcing half of them), changed the HRA completion incentive from a $10 gift card to a book of Forever postage stamps, and removed the C. Everett Koop endorsement letter from the mailing materials. A CMS "bridge letter" was also developed for the Year 1 beneficiaries and sent with a free calendar. Vendor B used the same recruitment process in Year 3 as in Year 2.

Tables 1 and 2 summarize aspects of each vendor's recruitment strategy.

Table 1: Vendor A Recruitment Summary

Vendor A
Year 1
Volume handled by mailing recruitment materials and conducting outbound awareness calls in six waves
Up to five awareness calls conducted to non-respondents (for each wave)
All awareness calls conducted with a live person
Live calls that provided more opportunity for opt-outs
Telematching service used to verify/obtain more addresses
Response rate 20% > vendor's most comparable target sample
Distinct opt-out (OO) categories: entire program OO, year-long OO, mid-year OO, coaching OO
Reasons for opting out collected
Year 2 Changes
Focused first on returning beneficiaries, then new beneficiaries; also sent CMS Bridge Letter
Year 3 Changes
Not Applicable

Table 2: Vendor B Recruitment Summary

Vendor B
Year 1
Recruitment materials sent out to entire sample; volume handled by outsourcing recruitment calls
Up to two awareness calls (three rounds of awareness calls, and a fourth to one Local sample)
50% awareness calls live, 50% automated (IVR)
Less opportunity for opt-outs with automated calls; opting-out designed to be active process
Telematching service not used
Response rate 20% < vendor's most comparable target sample
Participants choosing to discontinue coaching or self-directed mailings not considered "opt-outs"
Reasons for opting out not collected
Year 2 Changes
Materials sent out in four waves; CMS Bridge Letter and CMS calendar also sent
Local sample targeted first
Time between questionnaire completion and first coaching call reduced
Year 3 Changes
Not Applicable

2.3.2 HRA Questionnaires

Each vendor used previously developed HRA questionnaire tools to design customized HRAs for the SRRD. Each vendor's HRA addressed the following 17 modifiable risk factors required by CMS:

1. Physical inactivity/lack of exercise
2. Poor nutrition
3. Smoking/tobacco use
4. Excessive alcohol consumption
5. High blood pressure
6. High blood glucose
7. High total cholesterol
8. Being overweight/obese
9. Inappropriate use of clinical preventive services
10. Depression
11. High stress
12. Lack of general well-being
13. Burden of providing care giving
14. Social isolation
15. Lack of motor vehicle/home safety
16. Falls (preventable accidents)
17. Polypharmacy/medication issues.

Vendor A. Vendor A took a variety of already developed HRAs, matched them to the 17 CMS risk factors, and included standardized questions specific to the older adult population. Its HRA took into account all 17 plus an additional four (financial barriers, transportation barriers, independence level, and life quality). Vendor A conducted focus group and cognitive testing to ensure the questionnaire was understandable (aimed at a 6^{th} grade reading level) and made it as short as possible to maximize the response rate. Vendor A also had a Spanish-language version of the questionnaire available, containing the same questions as the English version. The English version indicated at the bottom of the form that the individual could request the questionnaire in Spanish.

Vendor A considered the HRA complete if 75 percent or more of the questionnaire was filled out. All completed HRAs were entered manually by vendor staff, and 25 percent were subjected to double data entry for quality assurance. No changes were made to the HRA for Year 2 of the demonstration.

Vendor B. The HRA developed by Vendor B captured all 17 of the CMS-indicated health risk factor areas. The English version was written at a 7th grade reading level and the Spanish version at a 6th grade level. The Spanish version also had fewer questions (32 versus 47).

Vendor B considered the HRA complete if 80 percent or more of the questionnaire was filled out and used an optical scanner to input the data from the completed HRAs.

The HRA used in Year 1 was changed slightly in Year 2 to remove questions related to family history because of the Genetic Information Nondiscrimination Act (GINA). These questions were replaced with new questions about skipping medications. The same HRA was used for both Year 3 and the Exit HRA.

Tables 3 and 4 summarize unique design and implementation features of each vendor's HRA.

Table 3: Vendor A HRA Summary

Vendor A
Year 1
HRA capture of all 17 CMS-indicated health risk areas plus four more
Successful enrollment defined as 75% of the HRA questionnaire filled out
Paper HRAs manually coded
Approximately 6% of HRAs completed online
HRA at 6th grade reading level
Same questions on Spanish version
Year 2 Changes
No changes made to HRA (to facilitate analysis across data points)
Year 3 Changes
Not Applicable

Table 4: Vendor B HRA Summary

Vendor B
Year 1
HRA capture of all 17 CMS-indicated health risk areas
Successful enrollment defined as 80% of the HRA instrument filled out
Paper HRAs optically scanned
Approximately 5% of HRAs completed online
English HSA at 7th grade reading level; Spanish version at 6th grade level
Fewer questions on Spanish version
Year 2 Changes
Family history question removed (due to GINA) and replaced with questions about skipping medications
Successful enrollment redefined as 75% of the HRA questionnaire filled out
Year 3 Changes
None

HRAs could be completed one of three ways—by paper, online, or telephonically (vendor staff would complete the data entry with the beneficiary over the phone). For each vendor, beneficiaries who successfully completed and returned an HRA became SRRD participants and were randomly assigned by CMS' implementation contractor to one of the three treatment arms.[30]

2.3.3 Risk Stratification and Intervention Services

The vendors tailored interventions to the SRRD participants based on the participant's Arm assignment. Each vendor also used its own proprietary risk stratification algorithms to stratify the SRRD participant population into low-risk, moderate-risk, or high-risk levels. According to the demonstration design, as noted, Arm 3 participants were only offered the HRA and generic health advice and did not receive any other type of intervention. Arm 2 (enhanced) participants generally received more interventions than Arm 1 (standard) participants; but within Arm 1 and Arm 2 the type and intensity of intervention services could vary based on the participant's risk category. Below, we outline each vendor's risk stratification process and describe the intervention services offered and how they were assigned.

Vendor A

Risk Stratification. Vendor A developed a new risk stratification algorithm for the demonstration. Although its development was guided by risk factors identified during work on a previous project, the algorithm itself was unique to the SRRD. The algorithm calculated

[30] For both vendors, the SRRD implementation contractor conducted the randomization of beneficiaries who returned completed HRAs, using a secure Simple Object Access Protocol (SOAP) randomization web service that assigned participants by beneficiary identifiers into one of the demonstration's three Intervention Arms.

weighted points for each of the CMS-defined risk categories, and the points determined whether the beneficiary was low, moderate, or high risk. The points (weighted based on the "impactability" of risk behavior) were totaled for each category and cut points defined where the participant was assigned to a particular risk category. The stratification algorithm placed participants with the heaviest burden of coaching/modifiable health risks into a high level of intervention, and those with less risk into moderate or low levels of intervention. To provide an enhanced intervention to Arm 2, the algorithm enabled the vendor to identify participants for additional services. Vendor A's risk stratification algorithm aimed to allocate approximately 10 percent, 30 percent, and 60 percent of participants into the low-risk, moderate-risk, or high-risk groups, respectively.

Program Enrollment and Intervention Services. Both vendors provided individualized HRA feedback reports that were closely reviewed with SRRD participants. For Vendor A, this took place during what was referred to as a "Lifestyle Management (LM)" or "Orientation" call conducted by a health advisor approximately five days after sending out the HRA feedback report. During this call, moderate- or high-risk participants could choose one or more (based on risk category) intervention program "focus areas"[31] as well as the mode of coaching (i.e., online or phone/mail-based[32] option). Vendor A developed a website for the SRRD that contained a variety of tools that eligible participants (Arm 1 and Arm 2) could use to assist in their health care management, including a health tracker, health calendar, information on allergies and medical conditions, a "my workouts" log, and a family health guide, among other features. For those who selected the online mode of intervention, "My Coach" access was provided in the website, including secure messaging with a health coach and access to focus areas. The low-risk participants in Arms 1 and 2 did not have the option to select an intervention mode and were only able to receive self-directed health improvement guides (one for Arm 1 and two for Arm 2). Although low-risk participants did not have the option to select online health coaching, they did have access to the website with the coaching option disabled.

For high- or moderate-risk participants, the LM call was followed by proactive calls from a health coach.[33] Low-risk participants could not receive proactive health coaching calls but were able to make "reactive" inbound calls to ask questions regarding the feedback report or educational materials. High- and moderate-risk participants in Arm 2 also had the option to participate in the Chronic Disease Self-Management Program (CDSMP) Webinar,[34] and high-risk Arm 2 participants were given the option of having a social worker or geriatric RN assessment for a community resource referral. Vendor A interventions available by Arm and risk category are depicted in Table 5.

[31] Arm 1 high-risk participants could select up to six focus areas; Arm 1 moderate-risk participants could select up to three; Arm 2 high- and moderate-risk participants could select up to six.

[32] All Arm 1 and 2 participants also received quarterly engagement mailers.

[33] Arm 1 moderate-risk participants got an average of three calls per year. Arm 1 high-risk participants and Arm 2 high- and moderate-risk participants got an average of six.

[34] Vendor A later discontinued the CDSMP Webinar option because of limited uptake.

Table 5: Vendor A Intervention Services by Arm[a]

Intervention Services / Materials	Arm 1 – Standard			Arm 2 – Enhanced			Arm 3
	Low	Mod	High	Low	Mod	High	All
Tailored Feedback Report w/Physician Summary Report	✓	✓	✓	✓	✓	✓	–
Health Advisor Orientation Call	✓	✓	✓	✓	✓	✓	–
Web Portal: Self-directed access (SRRD Action Guides, Family Health Guide, tracking tools, etc.)	✓	✓	✓	✓	✓	✓	–
Web-based Coaching: My Coach access in web portal (secured messaging with coach, web-based Empowered Focus areas) [Avg.] Number of coaching contacts][b]	–	✓ [3]	✓ [6]	–	✓ [6]	✓ [6]	–
Self-directed Health Improvement Guides [Avg. number of guides]	✓ [1]	–	–	✓ [2]	–	–	–
Empowered Health Coaching Guides (Hardcopy, including FYI Guides) [Avg. number of guides][c]	–	✓ [3]	✓ [6]	–	✓ [6]	✓ [6]	–
Proactive "how's it going" coach calls [Avg. number of calls]	–	✓ [3]	✓ [6]	–	✓ [6]	✓ [6]	–
Quarterly Engagement Mailers or Email	✓	✓	✓	✓	✓	✓	–
Help Line (technical)	✓	✓	✓	✓	✓	✓	✓
Chronic Disease Self-Management Program (CDSMP Webinar)[d]	–	–	–	–	✓	✓	–
Social Worker/Geriatric RN assessment for community resource referral	–	–	–	–	–	✓	–
Generic "Be Healthy" Letter (CMS-authored/approved)	–	–	–	–	–	–	✓

[a] Adapted from a table developed by Vendor A.
[b] Only for participants who selected web-based mode of coaching *instead* of phone/mail-based coaching.
[c] Only for participants who selected phone/mail-based coaching *instead* of web-based mode of coaching. Arm 1 Moderate Risk participants could select up to 3 focus areas; Arm 1 High Risk, and Arm 2 Moderate and High Risk participants could select up to 6 focus areas.
[d] In the enhanced Arm 2, the presence of a chronic condition determines eligibility for the online CDSMP workshop.

Vendor B

Risk Stratification. Once the initial randomization of the SRRD participants into one of the three treatment arms occurred, Vendor B used two predictive models it had already developed to assign participants into risk categories based on short- and long-term risk. Short-term risk was

based on the Utilization Prediction Model (UPM), a predictive modeling tool that projected total health services use for the participant in the following 12-month period. Long-term risk was based on the Lifestyle Score (LS), a predictive model that estimated avoidable health care costs based on demographic and health risk data collected in the HRA. Smoking and BMI data supplemented the predictive model in determining the risk category for each participant, and age and gender were also used for stratification. The short-term risk model generated a lifestyle score between 1 and 100, representing the participant's total controllable risk. Based on this score, 85 percent of SRRD participants were assigned into the high- or moderate-risk categories, and 15 percent into the low-risk category. Risk status determined eligibility for Vendor B's customized intervention programs, called the NextSteps® Lifestyle Behavior Change programs. Only moderate- and high-risk participants (considered "at-risk" participants) were eligible for the NextSteps programs; low-risk participants only received their HRA results and were not eligible for any intervention services. Vendor B's triaging did not distinguish between moderate- and high-risk participants.

Program Enrollment and Intervention Services. After completing the HRA, each participant received feedback including a lifestyle and disease risk score and a NextSteps program invitation, if eligible. A "state of change" model was used to help tailor the HRA feedback. Upon receiving the NextSteps invitation, the participant could actively enroll into a topic and a program delivery method. Participants who completed the HRA online would receive instant HRA feedback and an automatic invitation to NextSteps, which allowed the beneficiary to select and enroll directly into a program from the website. Otherwise, the vendor "bulk registered" eligible Arm 1 and Arm 2 participants into a recommended telephone program. Health coaches then placed a follow-up phone call to each participant for the intervention program, during which participants could elect to change the topic for enrollment, as well as the method of delivery.[35] The NextSteps programming was designed to address the following eight modifiable health behaviors: back care, blood pressure, cholesterol, nutrition, exercise, weight management, stress management, and tobacco cessation.

As summarized in Table 6, those in Arm 1 could choose a telephone-based, mail-based, or online-based mode of intervention. For the telephone-based intervention, each "at-risk" participant was matched to a personalized coaching intervention with a qualified health coach and could receive up to five coaching calls per year. The mail-based intervention option included a series of six monthly educational, motivational, interactive mailings to support behavior change. The online option included full access to the Healthy Living Program online suite, consisting of six health coaching sessions a week or more apart. Vendor B customized its existing online suite for the SRRD. The suite provided a variety of tools and interactive modules personalized for participants and allowed eligible participants (Arm 1) to set up health-related goals and to track progress towards meeting those goals. Among other features, the "health tools" section contained self-paced modules with interactive tools—including assessments, calculators, and quizzes that helped participants develop skills for managing asthma, diabetes, nutrition, weight, cholesterol, tobacco cessation, high blood pressure, heart health, fitness, and

[35] To be eligible for a program, the participant must have been assigned to the high- or moderate-risk category.

back care. In Vendor B's design, Arm 1 participants—regardless of risk level—could not receive the same or a higher level of resource-intensive services (e.g., number of coaching calls) than Arm 2 participants received.

Arm 2 participants could only participate in the telephone-based intervention option, consisting of up to 12 health coaching calls per year and at least one mailing per year based on a topic area triggered by their HRA results (see Table 6); however, the health coaching program was centered around the participant's needs rather than a set number of calls. Although Arm 2 participants could not receive access to the full online Healthy Living Program, they did have access to a health portal portion of the suite. Arm 2 participants could also participate in intervention programs such as Health Care Consumerism and Diabetes Management, which were not available to Arm 1 participants.

Table 6: Vendor B Intervention Services by Treatment Arm [a]

Intervention Materials/Services	Arm 1 – Standard	Arm 2 – Enhanced	Arm 3
Health Questionnaire (HQ)	HQ results provided by paper or online	HQ results provided by paper or online	–
Coaching Calls (phone program only)	Up to five per year	Up to 12 per year	–
Mailings (mail-based program only) [b]	Six customized mailings (one per month)	–	–
Vendor B Online Suite	Access to full online suite	Vendor B health portal only	–
Diabetes Management	–	Additional mailing available, if applicable	–
Health Care Consumerism	–	Additional mailing available	–
General Health Brochure	–	–	Brochure mailed instead of HQ results
Welcome mailing with the Healthy Approaches material	Sent to those who engaged with an intervention program[c]	Sent to those who engaged with an intervention program[c]	–

[a]Developed by the evaluation team based on information shared during the site visit.
[b]Beneficiaries could only be enrolled in one program; Arm 1 participants had the option of choosing a mail-based program *instead* of the phone-based program.
[c]Defined as having completed at least one call in a NextSteps program.

Key elements of each vendor's risk stratification and intervention process are shown in Tables 7 and 8.

Table 7: Vendor A Risk Stratification and Intervention Summary

Vendor A
Year 1
All 17 CMS-indicated risk areas taken into account plus four more
Stratification algorithm developed from "scratch" for the SRRD
All risk factors contributing to the risk level through weights; risk category determined by threshold values
Age and gender not factored into stratification
Distinguished between participants with different levels of risk: ~10% High risk ~30% Moderate risk ~60% Low risk
Some intervention received by 100% of Arm 1 and Arm 2 participants
Participants in a given Arm (1 or 2) triaged into different interventions based on risk stratum
Arm 2 participants eligible for all Arm 1 services and some additional services
Arm 1 (High): avg. six coaching calls Arm 1 (Moderate): avg. three calls Arm 2 (High and Moderate): avg. six calls
Readiness-to-change not included in Triage algorithm
All risk factors addressed in HRA feedback report but only 12 addressed graphically
Health advisors (orientation call) and health coaches
Chronic Disease Self-Management Program (CDSMP) Webinar offered to Arm 2 participants
Year 2 and 3 Changes
No changes

Table 8: Vendor B Risk Stratification and Intervention Summary

Vendor B
Year 1
8 of 17 CMS-indicated areas taken into account (based on vendor's existing model)
Existing predictive models used for stratification
Predictive model supplemented by smoking and BMI in determining risk category
Age and gender factored into stratification
Participants with different levels of risk minimally distinguished: ~85% High or Moderate risk ~15% Low risk
Some intervention received by 85% of Arm 1 and Arm 2 participants
All participants in a given Arm triaged into the same intervention (did not depend on risk stratum once in the 85%)
Arm 2 participants not allowed to access mail-based or web-based coaching interventions
Arm 1 (High and Moderate): up to five calls Arm 2 (High and Moderate): up to 12 calls
State of change, self-efficacy, and motivation included in triage algorithm
Life-style score addressed graphically in HAS feedback report
Health coaches only
Choice of Diabetes Management and Health Care Consumerism programs offered to Arm 2 participants
Year 2 and 3 Changes
No changes

2.4 Local Implementation

In the local demonstration component, participants in each vendor's two Local samples were referred to the ADRCs assigned to that vendor in the localities. The vendors and the local ADRCs began coordinating with each other prior to the SRRD implementation start date. Coordination activities included making key staff introductions, reviewing the purpose of the demonstration and proposed activities, and discussing the roles and responsibilities of each organization. The ADRCs and the vendors jointly developed memoranda of understanding to confirm each other's roles and responsibilities.

The general role of each ADRC was to review a list of local SRRD participants provided by the vendor on a regular basis and make initial outbound and follow-up telephone calls to inform them about the ADRC's services and available health risk reduction programs, resources, and services in the local community.[36] The referral lists from the vendors included contact and demographic information, as well as information about the specific health topics the participant chose to enroll in, which allowed ADRC staff to customize each phone call and offer referrals aligned with the participant's needs and goals. The ADRCs' existing data management systems

[36] For Vendor B, the initial plan was for the ADRC to accept "warm transfers" from the vendor (i.e., directly connecting the SRRD local participant to someone at the ADRC at the conclusion of a health coaching call); however, the process was discontinued.

were used to track, through case logs, all contacts and services provided to SRRD participants. Staff entered detailed notes into the system that provided a narrative summary of each call, including a plan of action and follow-up. Each ADRC was responsible for reporting back to the vendors on the outcomes of each contact—with cases considered closed if the individual could not be reached after multiple contact attempts, if the participant was reached but did not want a referral, or if the participant had successfully acted on a referral. The ADRCs had access to comprehensive databases on local community health programs and on resources and services SRRD participants could be referred to, which were updated on a regular basis.

Although the ADRCs had proven capabilities for providing information, referral, and assistance services to seniors in their communities, the design of the SRRD posed some unique challenges; and each ADRC reported that their initial understanding of how the program would unfold evolved as contacts with local SRRD participants were made. ADRC staff usually did not make outbound calls to their normal consumer base; rather, they were accustomed to taking inbound calls from a significantly older and frailer population in more immediate need of services. The ADRC staff reported that SRRD participants, for the most part, were healthier and more active than their usual clients and were largely uninterested in the services the ADRCs offered. For this reason, the ADRCs modified their outbound call scripts to move away from a "health counseling" approach and focus more on providing information about resources available in the community through newsletters and resource guides. Both vendors assisted their ADRCs with this process. For instance, Vendor B prepared an outbound call script for the ADRC staff to use, which included basic instructions on how to proceed if the SRRD participant was not interested in hearing about referrals (e.g., asking if the individual would like to be placed on the ADRC's mailing list). Vendor A held recurring three-way conference calls with its ADRCs to discuss how to address the lack of referral uptake, which included offering resource guides and bi-monthly newsletters instead of specific referrals. Despite these challenges, the ADRCs found that taking a proactive approach to engaging consumers in their communities was valuable in helping their staff develop new skills.

2.5 Conclusions and Lessons Learned

The SRRD was designed to test whether the success of well-structured health risk reduction programs in the private sector can be replicated by appropriately adapting those programs to the Medicare population to reduce health risk factors, improve health, and produce cost savings for the Medicare trust fund. Based on reviewing SRRD design documents, discussions with CMS and the SRRD implementation contractor, as well as site visits and interviews with vendors, we examined the processes each of the vendors followed to implement the demonstration. The following summarizes lessons learned from analyzing the implementation process.

Recruitment and Retention. Special attention had to be paid by the health management vendors to the design and branding of the SRRD recruitment materials to improve the credibility of the SRRD program to Medicare beneficiaries. Vendor staff reported that

beneficiaries had serious concerns about the legitimacy of the program and were hesitant to provide any personal information over the phone. Medicare beneficiaries are also accustomed to receiving and discarding unsolicited mail, which included SRRD recruitment materials. The beneficiaries included in the SRRD sample are heavily targeted by telemarketers, so strategies had to be developed to steer away from the use of rigid scripts by vendor staff conducting outbound awareness calls and focus on personalized messages instead. Consistent design for all communications, including an emphasis on using "Medicare branding," helped to increase trust and contributed to the successful recruitment results.

Use of a staggered/wave approach to recruitment proved to be a more effective strategy than sending out all recruitment materials at once. For Vendor B, shifting to the wave approach resulted in improved timing of awareness calls after the initial materials were sent, since the volume was more manageable. Vendors emphasized the importance of awareness calls as a strategy for increasing response and participation rates. In particular, one vendor was accustomed to working only with employer populations that communicated with their employees about participation, so the value of awareness calls for the SRRD was an important learning lesson.

In general, the vendors reported difficulty engaging the SRRD population. The vendors normally engage employee populations where significant awareness-building can occur at the worksite, which was not the case, of course, for the SRRD population. In addition, the vendors reported that not knowing which treatment arm the participant was assigned to was a recruitment barrier—preventing vendor staff from being able to explain to beneficiaries what specific services they would receive and what the benefits of participation would be.

Program Enrollment and Interventions. Some beneficiaries who selected the web-based mode of coaching were not able to successfully participate due to lack of familiarity and proficiency with the interactive web-based programs; for this reason one vendor modified the initial health coach scripting to include additional screening questions about beneficiaries' computer skills. For future risk reduction programs targeting the same type of population, additional attention could usefully be paid to the proportion of services offered through an online modality versus phone, because of the limited uptake of the online mode.

One vendor decided to switch from an "opt-in" to an "opt-out" model for program enrollment, because many participants were not voluntarily enrolling in its health improvement programs. The opt-out model proved successful because it automatically enrolled beneficiaries into a particular program based on their answers to the HRA—with participants able to change the program they were initially enrolled in during the first coaching contact.

Local Implementation. Although ADRCs were accustomed to fielding numerous calls from seniors in their communities and offer a wide range of health-related programming consistent with the SRRD objectives tailored to the needs of the local aging population, staff were not necessarily used to conducting outbound calls to individuals who had not previously contacted the ADRC for services on their own. ADRC staff often perceived these outbound calls as "cold

calls," which required careful scripting to make the SRRD participant aware that the call was related to their SRRD participation, and not telemarketing.

ADRC staff also reported that the characteristics of the local SRRD population impeded their ability to successfully connect participants to local community programs, services, and resources. Specifically, the SRRD participants appeared much healthier on average than the ADRCs' usual consumer base and were generally disinterested in taking advantage of the ADRC's information, referral, and assistance services. In one of Vendor B's local areas, the ADRC staff thought the SRRD participants were also wealthier on average than their usual consumer base, enabling them to have their own private resources and services already in place to manage their health.

3. SRRD PARTICIPATION

3.1 Introduction

In this chapter we examine beneficiaries' participation in the demonstration and their characteristics. In particular, we consider the following questions:

- Do participation rates differ across beneficiary characteristics such as demographics, chronic condition burden, and baseline Medicare spending and use?
- Does the length of time during which a beneficiary participates in the demonstration differ by treatment Arm?
- What are the characteristics of participating beneficiaries? How do participants' characteristics differ from those of the IG (i.e., the larger group of individuals invited to participate in the SRRD, from which participants were successfully recruited)?
- Do characteristics of beneficiaries in the refresh sample (beneficiaries added at the end of the first year of the demonstration) differ from those included in the original sample?

Summary of Findings

- SRRD participation rates were:
  - Higher among non-Hispanic whites compared with other race groups and among non-dual eligible individuals compared with dual eligible individuals,
  - Higher among individuals with cancer, osteoporosis, and arthritis and lower among individuals with Alzheimer's disease and stroke, and
  - Higher for the middle two quartiles of Medicare spenders than for the highest and lowest quartiles.
- Individuals assigned to the HRA-only Arm participated for a longer time than did participants assigned to treatment arms with follow-up service interventions.
- SRRD participants were:
  - Less likely to be members of a minority group and to be dual eligible, and
  - Less likely to have Alzheimer's disease and heart failure, and more likely to have arthritis and preventive screenings compared to the larger IG invited to be SRRD participants.
- For Vendor B, SRRD participants (compared to the larger IG invited to be SRRD participants) had higher Medicare expenditures in the baseline period.
- The refresh sample was similar to the original sample across most beneficiary characteristics during the baseline period.

We obtained program information from CMS and vendors to identify members of the IG and program participants. We then linked Medicare beneficiary administrative data for a baseline period before the start of the demonstration through its end. In this section, we provide a description of the data and our methodologies for addressing the research questions listed above and present our findings.

3.2 Data and Methodology

3.2.1 SRRD Design

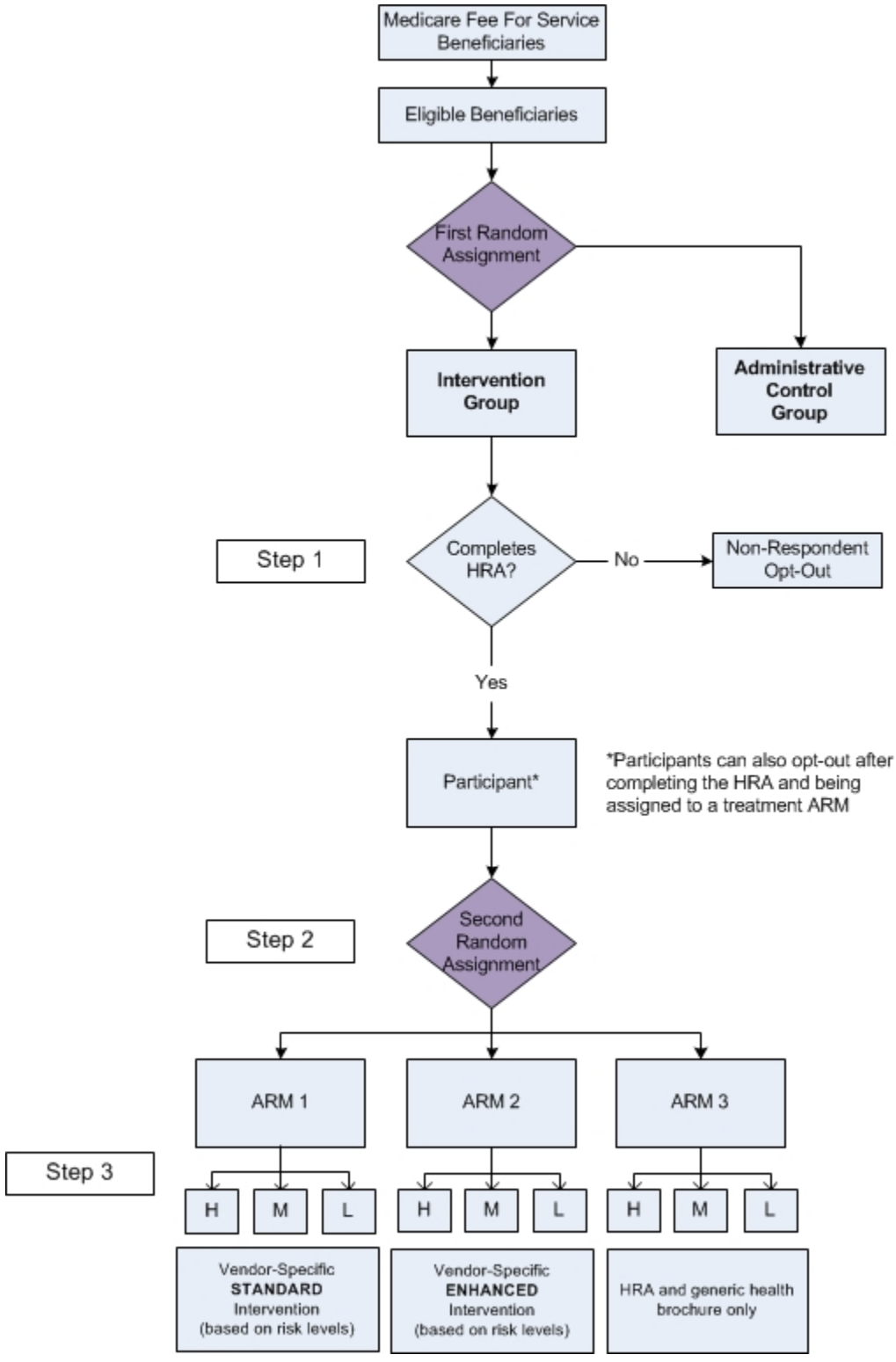
The SRRD demonstration had a RCT design with two rounds of random assignment. For each vendor, 40,000 eligible Medicare beneficiaries were randomly assigned to an Intervention Group (IG) or an Administrative Control Group (ACG) (20,000 to each group).[37] Each vendor then mailed a health risk assessment (HRA) packet to its assigned IG beneficiaries to recruit them for the study (i.e., invite them to participate). Figure 1 provides an overview of the SRRD design. Once a vendor received a completed HRA from a beneficiary, the beneficiary was considered successfully recruited into the study and became an SRRD participant (Step 1 from Figure 1). Participants were then randomized a second time into one of three intervention arms for each vendor (Step 2 from Figure 1):

- Arm 1 – Standard Treatment (HRA + standard tailored follow-up)
- Arm 2 – Enhanced Treatment (HRA + enhanced tailored follow-up)
- Arm 3 – HRA Only (HRA + generic health advice)

Arms 1 and 2 participants received varying levels of tailored intervention services aimed at reducing risk and improving health depending on the treatment arm to which they were randomly assigned and each beneficiary's HRA responses. Arm 3 participants received only a generic health mailing. The SRRD demonstration started May 1, 2009 and ended April 31, 2012.

[37] Vendors received a "refresh" sample consisting of an additional 11,800 individuals for each vendor at the end of the first year of the demonstration.

Figure 1. SRRD Study Design



Note: H = High Risk | M = Moderate Risk | L = Low Risk

3.2.2 Medicare Administrative Data

The Medicare administrative data used for this evaluation includes claims from Medicare Part A and B, demographic and enrollment files, and information on chronic conditions available from the Chronic Condition Warehouse (CCW). These files provide a rich source of beneficiary health information on the universe of Medicare Fee-for-Service (FFS) beneficiaries (approximately 40 million nationwide in 2008).

The Medicare Part A claims data consist of administrative FFS claims from institutional health care providers and include the following claim types: inpatient (from inpatient hospital providers for reimbursement of facility costs); skilled nursing facility; outpatient (from outpatient providers such as hospital outpatient departments, rural health clinics, renal dialysis facilities, outpatient rehabilitation facilities, and community mental health centers); hospice; and home health. The files contain variables such as diagnosis and procedure codes, dates of service, reimbursement amounts, and provider numbers.

The Medicare Part B claims data consist of administrative FFS claims from non-institutional health care providers. They include two claim types: carrier (from non-institutional providers such as physicians, physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers, and free-standing ambulatory surgical centers); and durable medical equipment regional carrier (from durable medical equipment suppliers). The files contain variables such as diagnosis and procedure codes, CMS Common Procedure Coding System codes, dates of service, reimbursement amounts, and provider numbers.

Demographic and enrollment information is available from the denominator file. This file contains state and county of residence codes, ZIP code, date of birth, date of death, sex, race, age, monthly entitlement indicators (A/B/C/D), reasons for entitlement, state buy-in indicators, and monthly managed care indicators (yes/no) for all Medicare beneficiaries enrolled and/or entitled to Medicare in a given year. The CCW Chronic Conditions File includes beneficiary flags for 21 chronic conditions generated by the presence of certain diagnosis codes during the previous one- or two-year look-back period, depending on the condition.

We obtained information from CMS on the two groups resulting from the first random assignment for each vendor (i.e., each vendor's IG and ACG). This information included identifiers of the beneficiaries originally assigned to each vendor ("original" sample) as well as of a "refresh" sample provided to vendors at the end of the Year 1 of the demonstration to assist in meeting recruiting targets. Vendors provided data on beneficiaries who were successfully recruited and participated in the demonstration. Using a linkable beneficiary identifier, we combined these beneficiaries with their Medicare administrative data.

3.3 Participation Rates

Beneficiaries who completed and submitted HRAs at the beginning of each demonstration year were considered participants for that year. Vendor A participated in the demonstration for two

years, so Vendor A's demonstration beneficiaries could participate in the program for up to two years. Vendor B participated for all three years of the demonstration, so Vendor B's demonstration beneficiaries could have participated for up to three years. In this section, we examine how participation rates varied by beneficiary characteristics. We also examine how length of participation varied by the treatment Arm to which the participant was assigned.

3.3.1 Rates of Participation

We calculated two types of participation rates: (1) the percentage of beneficiaries who participated in the program for at least one year ("any-year participants") and (2) the percentage of beneficiaries who re-enrolled ("multi-year participants"). We present the rates separately for each vendor and disaggregated by beneficiary characteristics that are policy and/or methodologically relevant:

- Gender
- Age
- Race/ethnicity
- Dual eligibility status
- Total Medicare expenditure
- Use of medical services covered under Medicare
- Presence of chronic conditions.

In addition, we used statistical inference to identify any differences in participation rates across beneficiary characteristics.

Note that the graphs in this section all feature 95 percent confidence interval bars to signify whether the differences are statistically significant or not. Only statistically significant differences are noted as differences in the text that follows.

Figure 2 displays participation rates for Vendor A across gender, age, race, and dual eligibility status for individuals who participated for at least one year (red circles) and for those who participated for multiple years (blue squares). There were no differences by gender or by age. Individuals who were Hispanic were least likely to participate—34.3 percent compared with 45.9 percent of non-Hispanic whites. Dual eligible individuals were less likely to participate—38.6 percent compared with 45.1 percent of non-dual eligibles. Any-year participation rates were higher than those for multi-year participation because beneficiaries participating in multiple years also participated in at least one year by definition. Vendor A's "any-year participation" rates were 44.1 percent among males and 44.6 percent among females, for example, compared with "multi-year participation" rates of 21.9 and 22.4 percent, respectively. The patterns for multi-year participation rates were generally similar to those for any-year participation rates.

Figure 2 provides similar information for Vendor B. There were no statistically significant differences by gender or age. However, participation was significantly higher for non-Hispanic whites than for other race/ethnicity categories, and for non-dual-eligible beneficiaries versus dual eligibles.

Figure 2: Vendor A – Rates of Any-Year Participation and Multi-Year Participation by Baseline Demographics

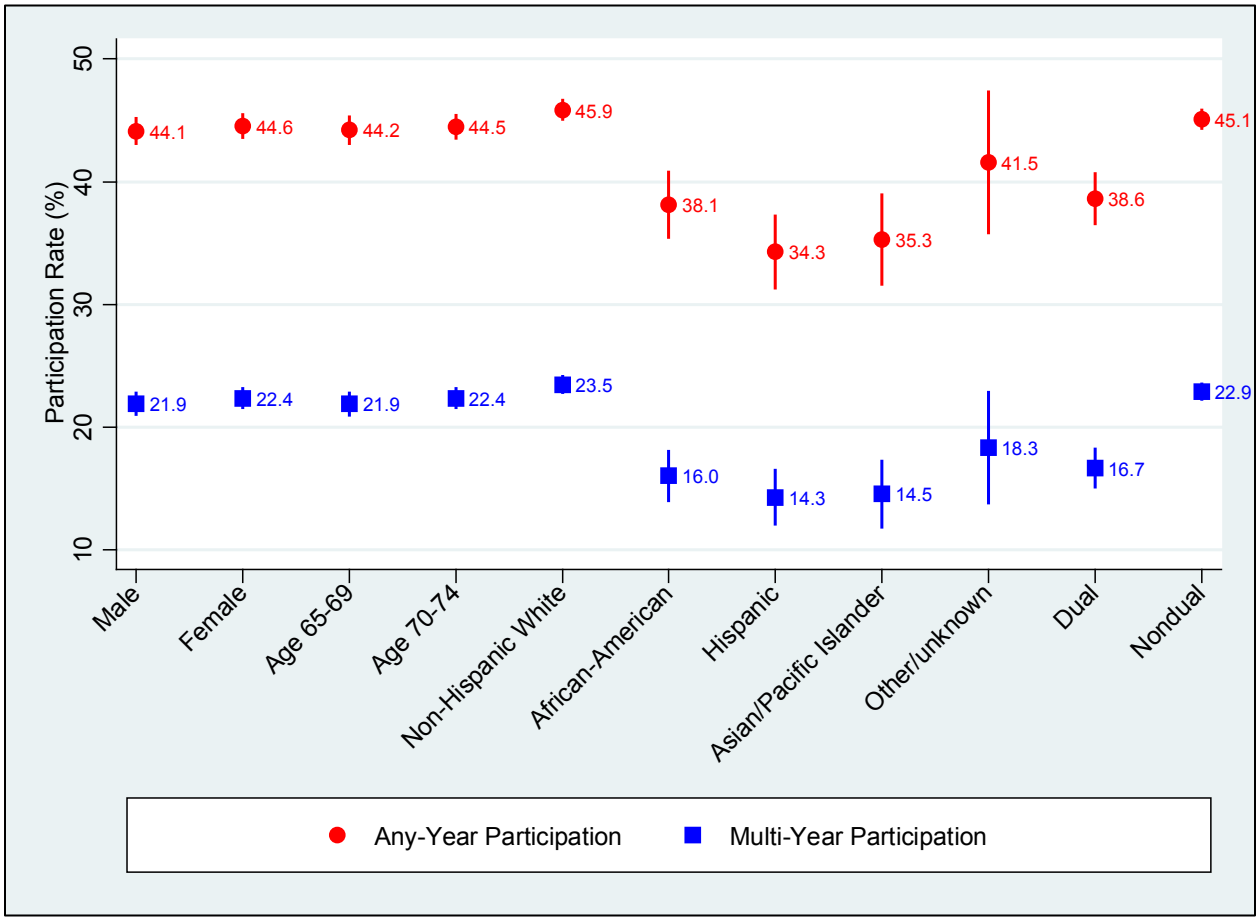
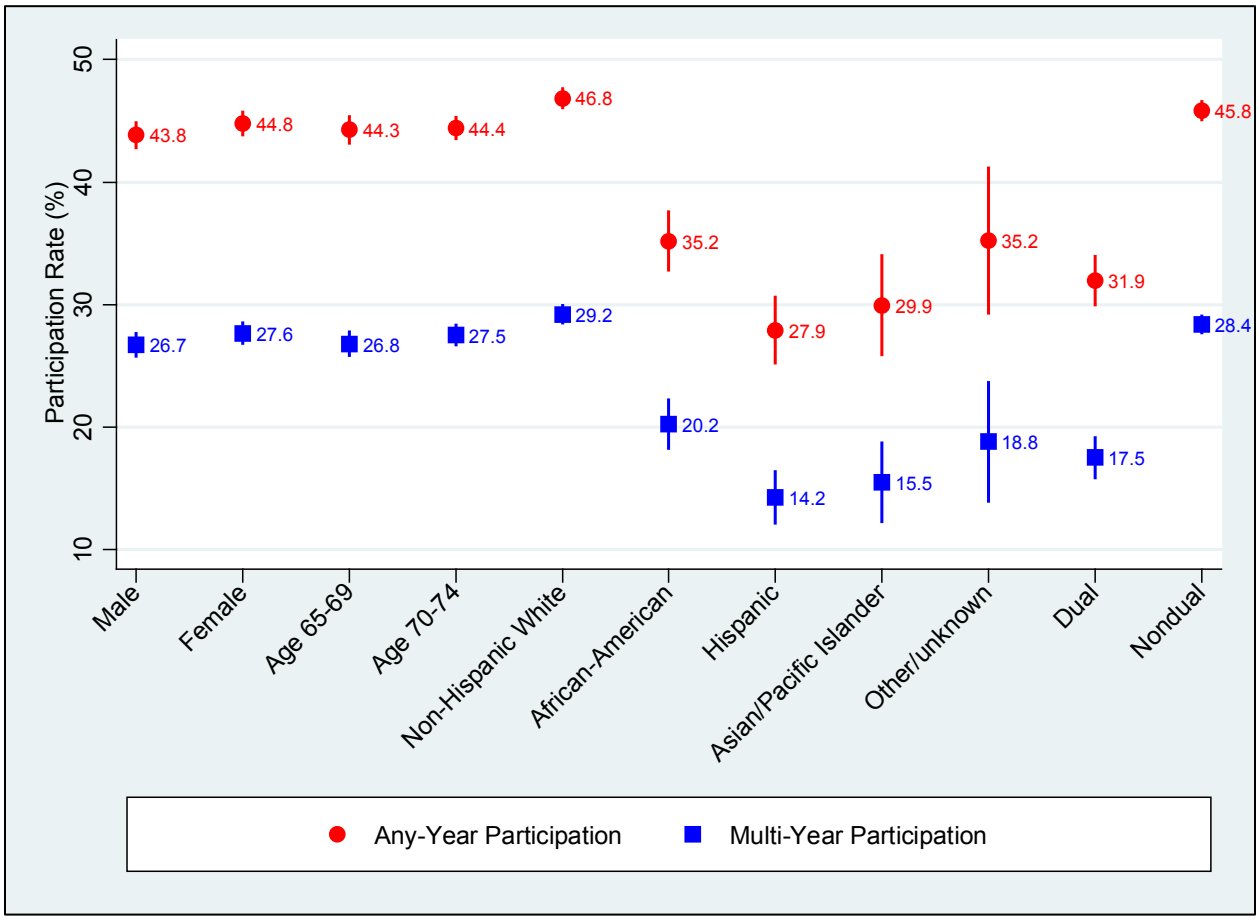


Figure 3: Vendor B – Rates of Any Participation and Multi-Year Participation by Baseline Demographics



Figures 4 and 5 provide information on participation by expenditure, health services use, and number of chronic conditions. Both vendors' participation profiles for expenditure quartiles follow an inverted-U pattern, with lowest participation among beneficiaries in the lowest quartile, next lowest by those in the highest quartile, and highest by those in the middle two quartiles (i.e., the middle 50 percent). The inverted-U pattern indicates that program take-up was highest among the population on which the intervention was likely to have the greatest effect.[38] In addition, for Vendor B, participants having expenditures in the top 5 percent had lower participation rates than the other 95 percent of enrollees. There were no statistically significant differences in participation rates for positive versus zero inpatient days or for two or more chronic conditions versus zero chronic conditions.

[38] The SRRD is likely to have the most impact among beneficiaries of average health status (i.e., those in the middle two total expenditure quartiles) rather than among the very sick (for whom no home-based intervention would be likely to prevent hospitalization) or the very healthy (for whom the hospitalization rate is very low). The study described in Wennberg, D.E., Marr, A., Lang, L., O'Malley, S., and Bennet, G. (2010), a randomized trial of a telephone care-management strategy, *New England Journal of Medicine*, 363(13), 1245-55, also uses this hypothesis.

Figure 4: Vendor A – Rates of Any Participation and Multi-Year Participation by Baseline Health Services Expenditure and Use

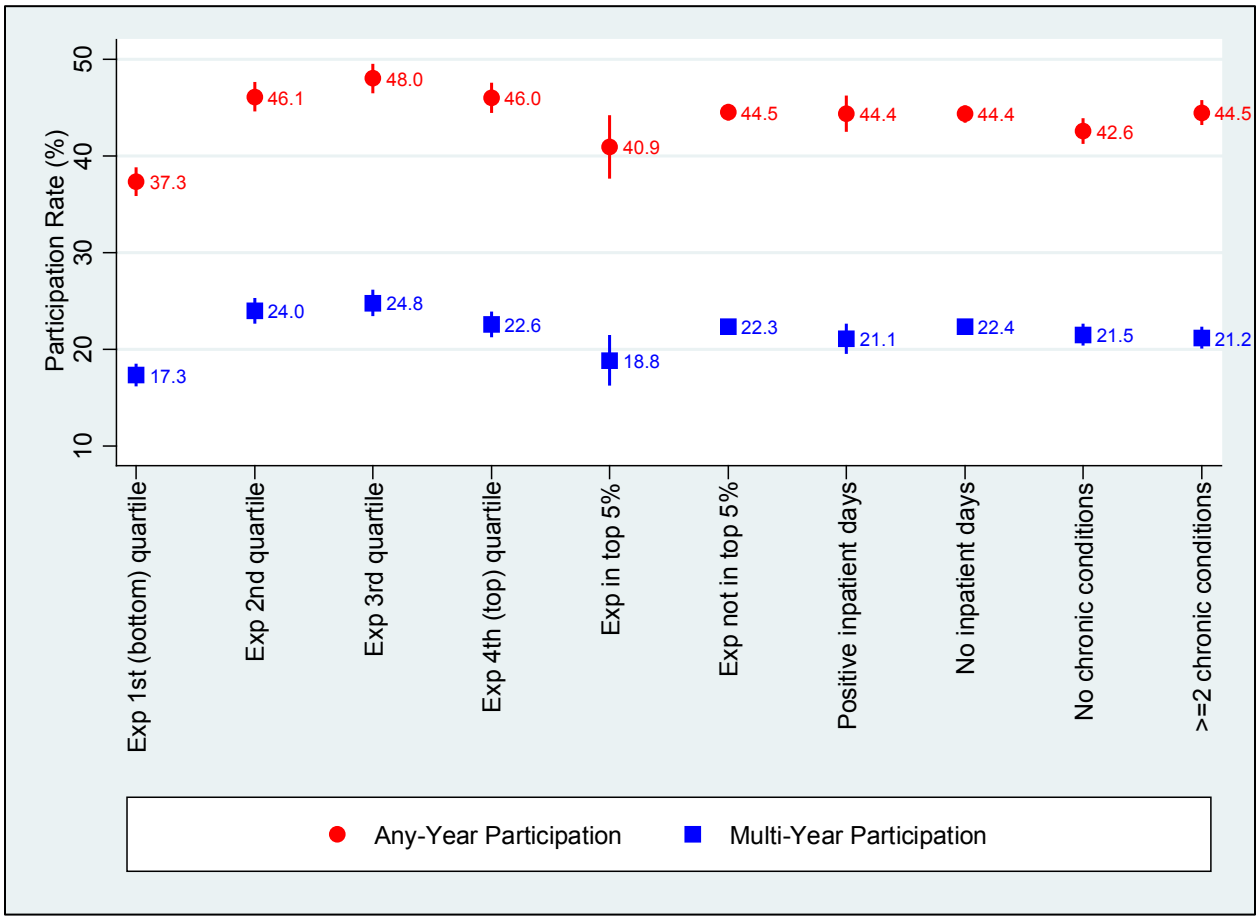
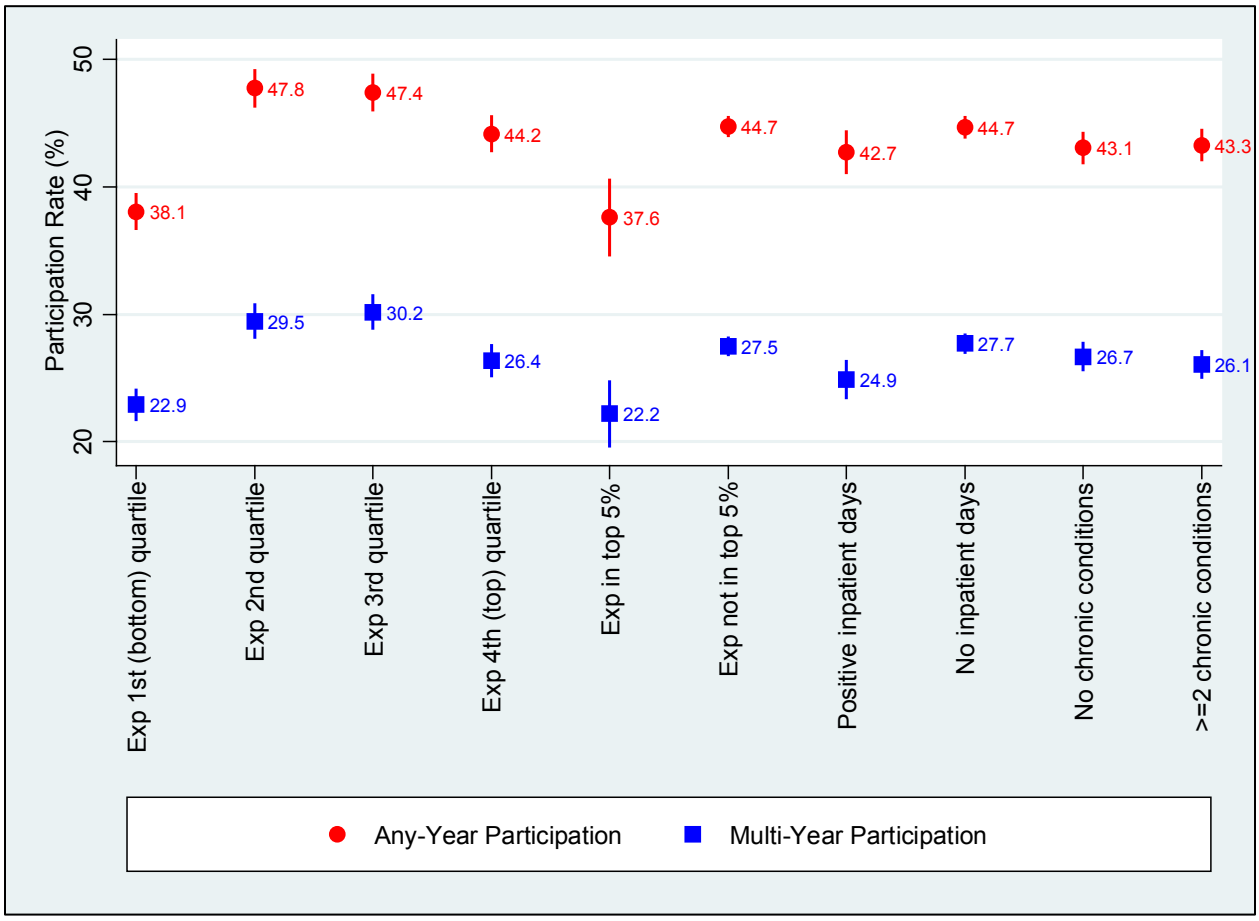


Figure 5: Vendor B – Rates of Any Participation and Multi-Year Participation by Baseline Health Services Expenditure and Use



Figures 6 and 7 examine participation rates for 11 specific chronic conditions. For both vendors, participation rates were highest among beneficiaries with cancer, osteoporosis, and arthritis, and lowest among those with Alzheimer's disease and stroke. Low participation among Alzheimer's and stroke patients was to be expected, since the demonstration required that beneficiaries have sufficient mental capacity to engage in the SRRD.

Figure 6: Vendor A – Rates of Any Participation and Multi-Year Participation by Baseline Chronic Condition

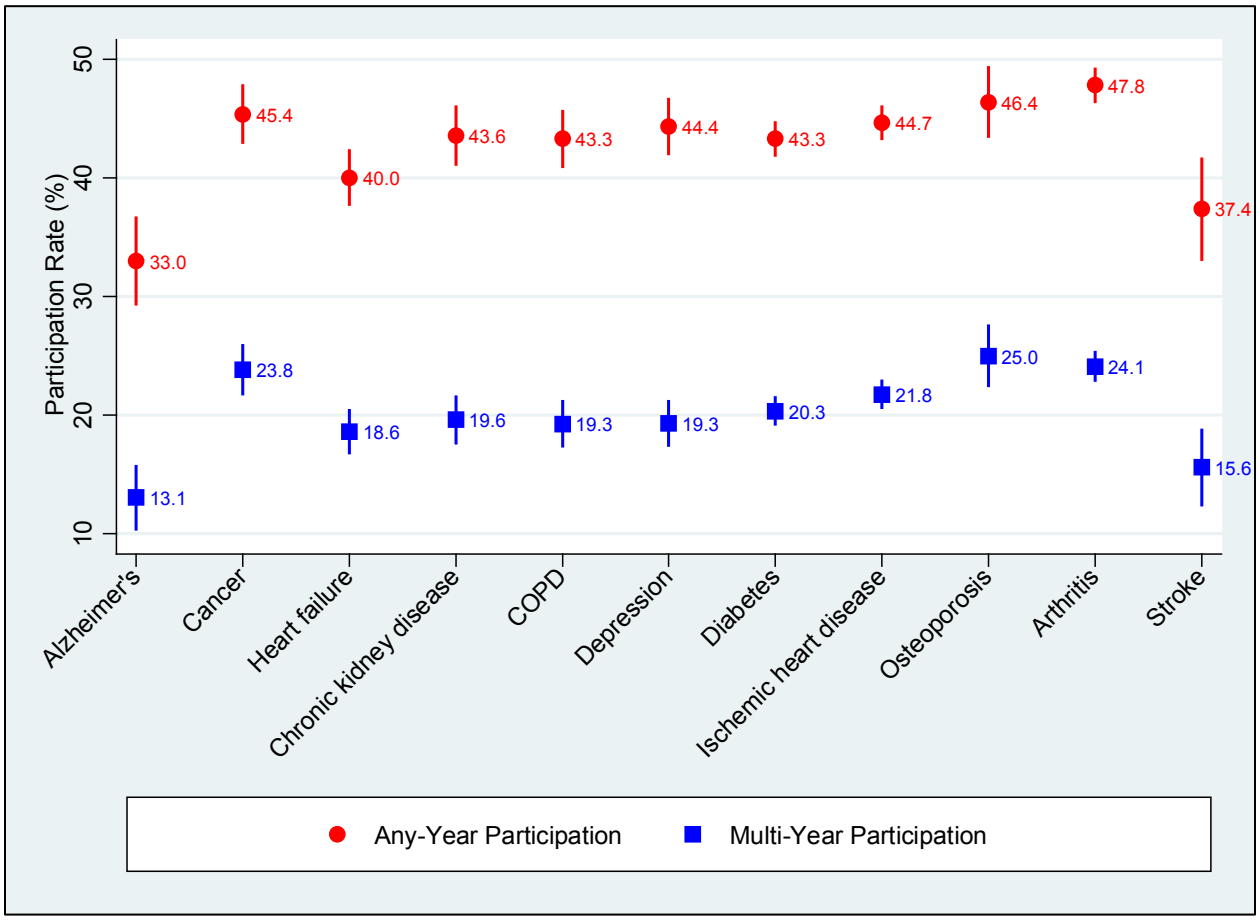
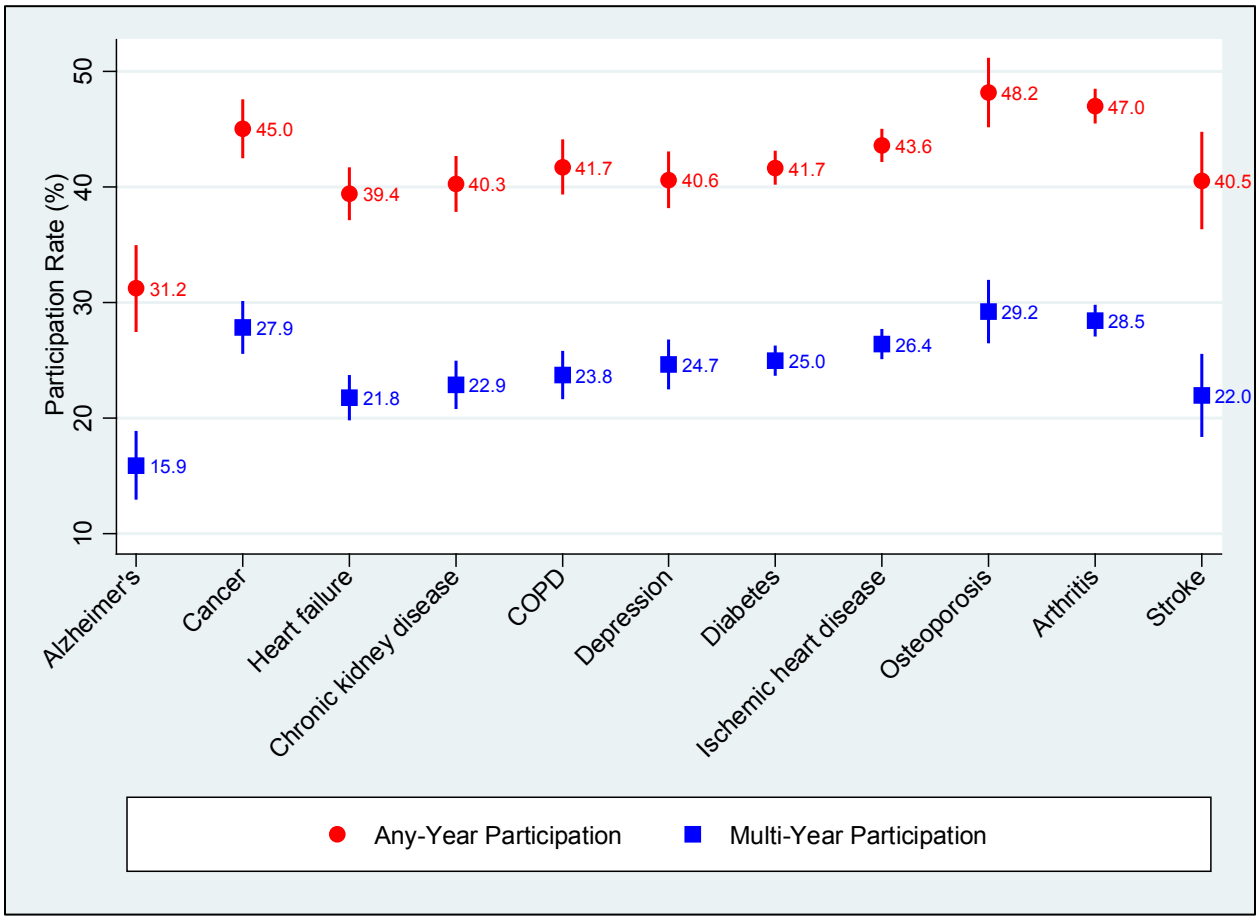


Figure 7: Vendor B – Rates of Any Participation and Multi-Year Participation by Baseline Chronic Condition



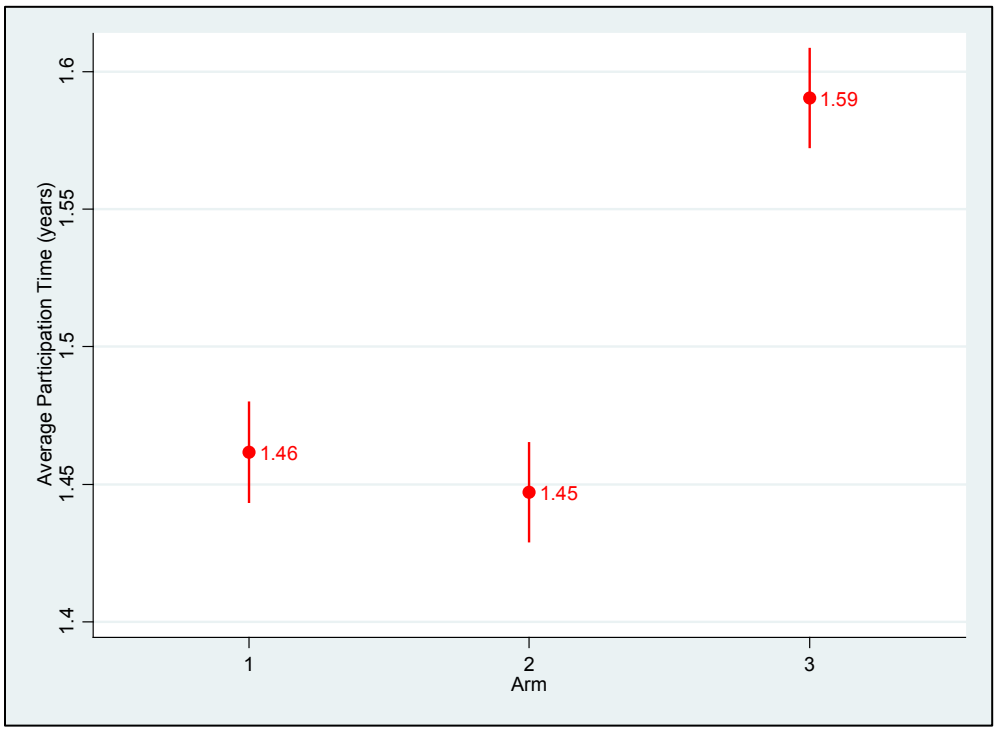
3.3.2 Average Length of Participation

In this section, we report on the length of time participants remained in the demonstration by calculating mean participation time in years for participants in each treatment Arm and separately for each vendor. We used statistical inference to determine whether there were differences in participation time across arms.

The SRRD was implemented on May 1, 2009 and lasted through April 30, 2012, with Vendor B continuing the demonstration throughout the three-year period. Vendor A discontinued participation after the end of the Year 2 (April 30, 2011). In our analyses, we use "post-implementation period" to refer to the time from April 2009 through the end of SRRD implementation (April 2011 for one vendor; April 2012 for the other). However, since the period for which a particular individual participated may also differ within vendors—some individuals only chose to participate for one year, while others re-enrolled for two or three years—we examined the average length of participation by treatment Arm.

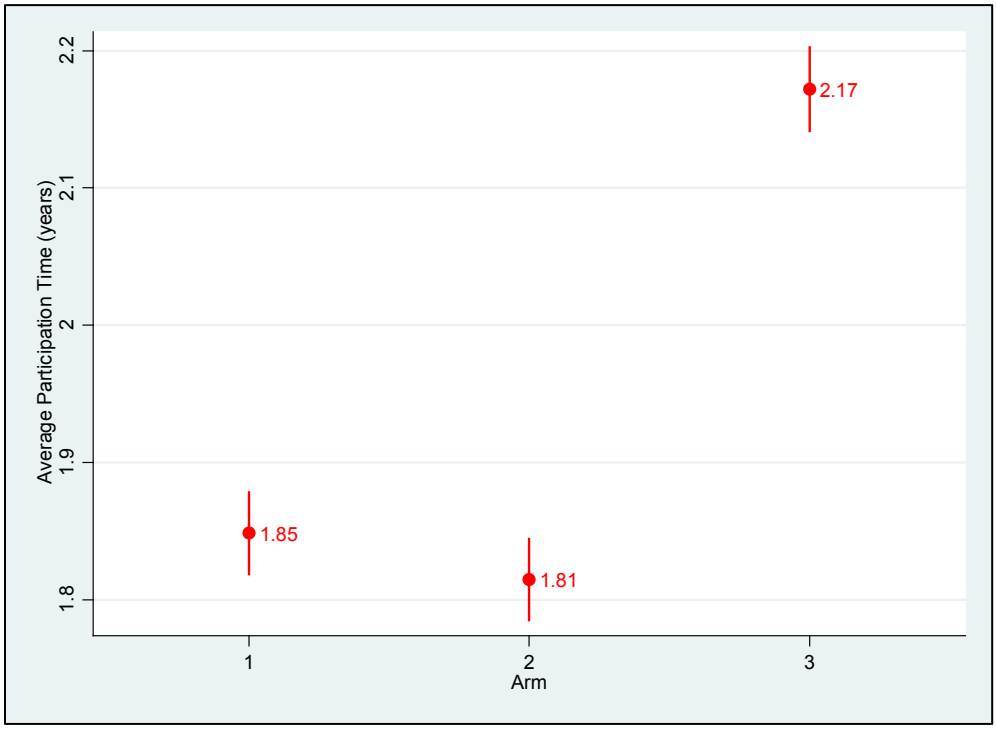
Figure 8 displays the average length of participation, including 95 percent confidence interval bars, by treatment Arm for Vendor A. The average length of time individuals participated was 1.46 years for Arm 1, 1.45 years for Arm 2, and 1.59 years for Arm 3. For Vendor B, the average length of time individuals participated was 1.85 years for Arm 1, 1.81 years for Arm 2, and 2.17 years for Arm 3 (Figure 9). For both vendors, average length of participation was statistically indistinguishable for Arms 1 and 2, but Arm 3 participants remained in the demonstration longer than did the other two arms. Arms 1 and 2 may have had lower retention if (1) beneficiaries found participating in the intervention burdensome or (2) the benefits of the intervention accrued quickly (so that wellness goals were achieved quickly) and participants considered continued enrollment to have small additional benefit.

Figure 8: Average Participation Time by Arm – Vendor A



Note: Vendor A had a total of 8,873 participants from the pool of original IG beneficiaries. These participants were evenly split among the three arms: Arm 1 - 2,957; Arm 2 - 2,966; Arm 3 – 2,950.

Figure 9: Average Participation Time by Arm – Vendor B



Note: Vendor B had a total of 8,870 participants from the pool of original IG beneficiaries. These participants were evenly split among the three arms: Arm 1 - 2,958; Arm 2 - 2,956; Arm 3 – 2,956.

3.4 Beneficiary Baseline Characteristics

In this section, we compare the characteristics of beneficiaries participating in the demonstration to those of the IG. We provide summary statistics for each of the beneficiary characteristics listed above for the IG, any-year participants, and multi-year participants. We used statistical inference to determine whether characteristics differed among these groups.

Table 1 presents information for Vendor A on the characteristics of beneficiaries participating in the SRRD and statistical tests of whether participants differed from the IG. Column 2 of Table 1 shows baseline beneficiary characteristics for Vendor A's IG. Column 3 shows baseline beneficiary characteristics for those individuals from the IG who participated in the demonstration. Column 4 shows the statistical significance of the differences between the two groups. Column 5 provides baseline beneficiary characteristics of participants who re-enrolled (i.e., participated for more than one year). Column 6 shows the statistical significance of differences between the baseline characteristics of beneficiaries participating in the demonstration for multiple years and those of beneficiaries participating in any year (from Column 3). Table 2 provides this information for Vendor B. Both tables pool Local and National samples and include all participants regardless of treatment arm assignment.

Both vendors' any-year participants were more likely to be non-Hispanic whites and less likely to be members of a minority group or dual eligible. For Vendor B, the proportion of non-Hispanic white participants was 4.8 percentage points higher than the proportion of non-Hispanic white beneficiaries in the IG. The corresponding figure for Vendor A was 2.8 percentage points. Similarly, for both vendors, multi-year participants are more likely to be non-Hispanic whites than are beneficiaries in the IG.

In addition to the race/ethnicity profiles of each vendor's participants, there were statistically significant differences between participants and the IG in (1) the prevalence of certain chronic conditions and (2) preventive screenings. Any-year participants were less likely to have Alzheimer's and heart failure and more likely to have arthritis and all three of the preventive screenings we examined (colorectal cancer, breast cancer, and cardiovascular). Vendor A's any-year participants were less likely to have had a stroke or TIA. Vendor B's any-year participants were less likely to have a number of chronic conditions (including kidney disease, depression, and diabetes) and less likely to have osteoporosis. Differences between the any-year and multi-year participant groups were generally not statistically significant for chronic conditions prevalence, but multi-year participants were more likely to have had some preventive screening than were any-year participants (breast cancer screening for both vendors and cardiovascular screening for Vendor A only).

Tables 1 and 2 also examine Medicare expenditures at baseline. From Column 6, it is clear that multi-year participants were similar to any-year participants in baseline expenditures. From Column 4, we find that Vendor A's participants and IG beneficiaries had similar total Medicare, inpatient, physician, and outpatient payments. Vendor B's participants had lower total Medicare and inpatient payments compared to IG beneficiaries, both statistically significant at the 5 percent level. However, this appears to be driven by differences in the vendors' IGs rather than differences in the utilization for each vendor's participants. In particular, Vendor B's IG had total Medicare payments of $6,148 per beneficiary and inpatient payments of $2,273 per beneficiary. These figures are $415 and $197 higher than the figures for Vendor A's IG total and inpatient payments, respectively. However, the vendors' participants had similar expenditures: Vendor B's participants' total and inpatient payments were only $2 lower and $3 lower than those of Vendor A's. This indicates that Vendor B's IG had relatively high baseline expenditures, but baseline expenditures for participants were similar across vendors.

Table 1. Vendor A - Participant Characteristics at Baseline

Baseline beneficiary characteristics[a]	IG	Any-Year Participants[b]	Diff.[c]	Multi-Year Participants	Diff.[c]
(1)	(2)	(3)	(4)=(3-(2)	(5)	(6)=(5)-(3)
	N=25,995	N=10,299		N=4,475	
Demographic characteristics					
Female	55.7%	55.7%	0	56.0%	0.3
Age (as of 5/1/09)	70.7	70.7	0.0	70.7	0.0
Race (RTI race code)					
Non-Hispanic White	83.8%	86.7%	2.8***	88.8%	2.1**
African American	6.5%	5.7%	-0.9**	4.7%	-1.0**
Hispanic	5.0%	3.8%	-1.2***	3.2%	-0.6*
Asian/Pacific Islander	3.3%	2.6%	-0.7**	2.2%	-0.4
Medicaid/Medicare dual eligibility	11.3%	9.6%	-1.6***	8.5%	-1.2**
Medicare use					
Total Medicare payment	$5,733	$5,690	-43	$5,469	-221
Inpatient payment	$2,076	$1,997	-80	$1,864	-13
Physician payment	$1,216	$1,278	62	$1,251	-28
Outpatient payment	$1,014	$1,023	9	$1,027	4
Chronic conditions[d]					
Alzheimer's/related disorders/senile dementia	3.0%	2.3%	-0.8**	1.9%	-0.4
Cancer	8.3%	8.4%	0.2	9.0%	0.5
Heart failure	8.9%	8.1%	-0.8**	7.4%	-0.6
Chronic kidney disease	8.1%	8.0%	-0.1	7.2%	-0.8
COPD	8.6%	8.4%	-0.2	7.6%	-0.8
Depression	8.7%	8.6%	-0.2	7.8%	-0.8
Diabetes	25.5%	24.8%	-0.7	23.3%	-1.6*
Ischemic heart disease	25.7%	26.0%	0.2	25.4%	-0.6
Osteoporosis	5.9%	6.1%	0.1	6.6%	0.5
Rheumatoid/osteoarthritis	25.0%	26.9%	1.9**	26.7%	-0.2
Stroke/TIA	2.5%	2.0%	-0.4**	1.7%	-0.3
Preventive screening use					
Colorectal CA screening	22.1%	24.3%	2.3***	24.9%	0.6
Breast cancer screening	58.5%	66.7%	8.2***	70.5%	3.8**
Cardiovascular screening	68.3%	72.7%	4.4***	75.0%	2.3**

[a]Characteristics are annualized figures based on information from the baseline period, January 1, 2008 to April 30, 2009, unless indicated otherwise.

[b] Any-year participants filled out an HRA at the beginning of Years 1, 2, and/or 3 of the demonstration.

[c] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

[d] Chronic conditions were measured at mid-year 2009.

Table 2. Vendor B - Participant Characteristics at Baseline

Baseline beneficiary characteristics[a]	IG	Any-Year Participants[b]	Diff.[c]	Multi-Year Participants	Diff.[c]
(1)	(2)	(3)	(4)=(3)-(2)	(5)	(6)=(5)-(3)
	N=26,000	N=9,882		N=6,053	
Demographic characteristics					
Female	55.0%	55.5%	0.5	55.7%	0.3
Age (as of 5/1/09)	70.7	70.8	0.01	70.8	0.0
Race (RTI race code)					
Non-Hispanic White	83.0%	87.8%	4.8***	89.1%	1.3**
African American	7.9%	6.2%	-1.7***	5.9%	-0.3
Hispanic	5.4%	3.4%	-2.1***	2.8%	-0.6*
Asian/Pacific Islander	2.4%	1.7%	-0.8***	1.4%	-0.3
Medicaid/Medicare dual eligibility	10.7%	7.6%	-3.1***	6.8%	-0.8*
Medicare use					
Total Medicare payment	$6,148	$5,688	-460**	$5,489	-199
Inpatient payment	$2,273	$1,994	-279**	$1,937	-57
Physician payment	$1,348	$1,358	10	$1,331	-27
Outpatient payment	$980	$995	15	$938	-58
Chronic conditions[d]					
Alzheimer's/related disorders/senile dementia	3.2%	2.2%	-1.0***	1.9%	-0.4
Cancer	8.3%	8.4%	0.1	8.5%	0.1
Heart failure	9.7%	8.8%	-1.0**	8.0%	-0.8*
Chronic kidney disease	8.9%	7.9%	-0.9**	7.5%	-0.4
COPD	9.0%	8.6%	-0.4	8.0%	-0.6
Depression	8.6%	7.9%	-0.7*	7.8%	-0.1
Diabetes	26.2%	24.4%	-1.8**	23.8%	-0.6
Ischemic heart disease	27.3%	26.7%	-0.5	26.3%	-0.5
Osteoporosis	5.8%	6.4%	0.6*	6.4%	0.0
Rheumatoid/osteoarthritis	25.5%	27.0%	1.5**	26.5%	-0.5
Stroke/TIA	2.8%	2.5%	-0.2	2.2%	-0.4
Preventive screening use					
Colorectal CA screening	22.0%	24.8%	2.7***	25.2%	0.5
Breast cancer screening	57.3%	66.7%	9.3***	69.2%	2.6**
Cardiovascular screening	68.0%	72.9%	4.9***	74.3%	1.5

[a] Characteristics are annualized figures based on information from the baseline period, January 1, 2008 to April 30, 2009, unless indicated otherwise.

[b] Any-year participants filled out an HRA at the beginning of Years 1, 2, and/or 3 of the demonstration.

[c] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

[d] Chronic conditions were measured at mid-year 2009.

3.5 Multivariate Regression Analyses of Participation

In this section, we address the following research questions:

1. Which beneficiary characteristics were associated with participation in the intervention?
2. Which beneficiary characteristics were associated with re-enrollment (i.e., participation over multiple years)? Did treatment Arm assignment affect whether or not an individual re-enrolled?
3. Did beneficiaries in the "refresh" sample provided at the end of Year 1 to assist vendors in reaching recruiting goals differ from the original sample?

To address these questions we used multivariate regression models, which allowed us to examine the effect of multiple characteristics simultaneously. We first provide detail on the methodologies used and then summarize the results.

3.5.1 Methodology

We estimated linear probability models separately for each vendor for the likelihood of:

- Any participation
- Participation in multiple years
- Being in the refresh sample.

Each of these outcomes is described below.

1. *Any participation (at least one HRA):* The outcome of interest is an indicator for any participation. For each vendor, the sample consists of all beneficiaries in the vendor's IG. The following control variables are included:
    - Age
    - Indicator for female
    - Indicator for nonwhite race
    - Indicator for dual status
    - Total Medicare expenditures during the baseline period
    - Indicator for any inpatient claims during the baseline period
    - Community risk score from the Medicare Part C risk adjustment files
    - Indicators for 11 chronic conditions (Alzheimer's disease, chronic kidney disease, chronic obstructive pulmonary disease, heart failure, diabetes, ischemic heart disease, depression, osteoporosis, arthritis, stroke/transient ischemic attack (TIA), and cancer).

2. *Multiple year participation (re-enrollment):* The outcome of interest is an indicator for whether the individual participated for more than one year. For each vendor, the sample consists of individuals who participated for at least one year (any participation). Control variables include those listed above plus the following:
    - Indicator for the treatment Arm to which the individual was assigned
    - Indicator for whether the individual rated his/her health status as poor[39]
    - Overall risk level based on HRA responses.[40]
3. *Refresh sample:* The outcome of interest is an indicator for whether the beneficiary was a member of the original sample or the refresh sample. We performed this analysis for all demonstration beneficiaries (IG and ACG members) as well as separately for those who participated for at least one year. The latter analysis allowed us to test whether members of the refresh sample who chose to participate in the demonstration were different from members of the original sample who chose to participate. For the full sample (IG and ACG individuals) we include the control variables listed in #1 above. For the participant samples, we include the control variables listed in #1 and #2 above.

3.5.2 Results

Tables 3 and 4 provide an overview of the results of these regressions. Each cell of the table represents a regression coefficient. Empty cells indicate that the parameter estimate was not statistically significant (p-value greater than 0.05) and the symbols "-" and "+" indicate that the estimate was statistically significant (p-value less than 0.05) and negative or positive, respectively.

As shown, nonwhites and individuals with Alzheimer's disease were less likely to have participated at all or over multiple years, and beneficiaries with arthritis were more likely to participate, though no more likely to participate for multiple years. Also, dual eligibles were less likely than non-duals to have participated at all. Participants with higher HRA risk levels and those assigned to treatment Arms 1 or 2 were less likely to participate for multiple years compared to individuals assigned to Arm 3; this is consistent with Figures 8 and 9, which indicate that Arm 3 participants participated for longer than Arm 1 or Arm 2 participants.

The analysis of inclusion in the original sample of beneficiaries suggests that the original members of the IG and ACG had higher total Medicare expenditures in the baseline period and were less likely to have arthritis than the refresh sample. However, most beneficiary characteristics were not consistently associated with inclusion in the original versus refresh sample. Importantly, the probability of being a member of the original sample is not associated with assignment to Arms 1 or 2 versus Arm 3.

[39] Self-reported health status was obtained from participants' HRA data.
[40] Overall risk was constructed by vendors using responses to the HRAs.

Table 3. Overview of Participation Analysis Regressions – Vendor A

	Any Participation	Multiple Year Participation	Refresh Sample (All)	Refresh Sample (Participants only)
Demographic characteristics				
Female				
Age	-			
Nonwhite	-	-		
Dual eligible	-	+		-
Medicare use				
Total Medicare payment			+	
Any inpatient claims				
Chronic conditions/health status				
Alzheimer's	-	-		-
Cancer				
Heart failure	-			
Chronic kidney disease				
COPD				
Depression				
Diabetes				
Ischemic heart disease				
Osteoporosis			-	
Rheumatoid/osteoarthritis	+		-	
Stroke/TIA	-		+	
HCC community risk score				
HRA Risk Level	NA	-	NA	
Poor self-rated health status	NA	-	NA	
Treatment arm				
Arm 1 or 2	NA	-	NA	

Table 4. Overview of Participation Analysis Regressions – Vendor B

	Any Participation	Multiple Year Participation	Refresh Sample (All)	Refresh Sample (Participants only)
Demographic characteristics				
Female				
Age		+		
Nonwhite	-	-		
Dual eligible	-			
Medicare use				
Total Medicare payment		+	+	
Any inpatient claims				
Chronic conditions/health status				
Alzheimer's	-			
Cancer				
Heart failure				
Chronic kidney disease				
COPD				
Depression	-			
Diabetes				
Ischemic heart disease		+		
Osteoporosis	+			
Rheumatoid/osteoarthritis	+		-	
Stroke/TIA				-
HCC community risk score		-		
HRA Risk Level	NA	-	NA	
Poor self-rated health status	NA		NA	
Treatment arm				
Arm 1 or 2	NA	-	NA	

Table 5 presents the details of the participation regression results for Vendor A. Columns 2 and 3 report the results of an indicator for any participation regressed on beneficiary characteristics. Columns 4 and 5 present the results of multiple year participation (or re-enrollment) on beneficiary characteristics and treatment Arm assignment. The table shows that for Vendor A, a one-year increase in age was associated with a 0.3 percentage point decrease in the likelihood of participating. Nonwhite beneficiaries were 7 percentage points less likely to participate than white beneficiaries and, nonwhite beneficiaries who participated for at least one year were about 5 percentage points less likely to re-enroll compared to white beneficiaries who participated for least one year. Dual eligibles were 3 percentage points less likely to participate, and for Vendor A, duals were less likely to re-enroll. Total Medicare payment and inpatient service use was not associated with participation for Vendor A.

We found lower participation among individuals with Alzheimer's (approximately 9 percentage points, p-value = 0.000) and higher participation among beneficiaries with arthritis (approximately 4 percentage points, p-value = 0.000). Higher overall risk levels as measured by

HRA responses were associated with decreased likelihood of participation in multiple years (p-value = 0.000).[41] Also, rating oneself as having poor health was associated with a 10 percentage point decrease in the likelihood of participating over multiple years for Vendor A (p-value = 0.000). Assignment to Arms 1 or 2 was negatively associated with participation in multiple years.

Table 5. Vendor A - Participation Regression Results

	Any Participation		Multi-Year Participation	
	Coefficient	P-value	Coefficient	P-value
(1)	(2)	(3)	(4)	(5)
Female	-0.002	0.791	0.010	0.403
Age	-0.003**	0.016	0.002	0.406
Nonwhite	-0.070***	0.000	-0.051***	0.005
Dual eligible	-0.034***	0.001	0.058***	0.008
Total Medicare payment	0.000	0.903	0.000	0.260
Any inpatient claims	-0.002	0.853	0.007	0.716
Alzheimer's	-0.088***	0.000	-0.105***	0.005
Cancer	0.004	0.734	0.038*	0.062
Heart failure	-0.042***	0.001	0.029	0.236
Chronic kidney disease	0.010	0.425	-0.038	0.102
COPD	-0.009	0.444	-0.018	0.430
Depression	-0.004	0.756	-0.035*	0.087
Diabetes	-0.005	0.494	0.014	0.340
Ischemic heart disease	0.012	0.119	0.005	0.744
Osteoporosis	0.009	0.482	0.035	0.139
Rheumatoid/osteoarthritis	0.045***	0.000	0.017	0.191
Stroke/TIA	-0.055***	0.007	-0.050	0.211
HCC community risk score	0.007	0.227	0.014	0.219
HRA Risk Level			-0.019***	0.000
Poor self-rated health status			-0.101***	0.000
Arm 1			-0.154***	0.000
Arm 2			-0.170***	0.000
N	25978		7304	
R-sq	0.008		0.050	

Statistical significance is shown at the 1% (***), 5% (**) and 10% (*) levels

Table 6 presents the results for Vendor B. Age was not associated with changes in likelihood of participation, but older beneficiaries who did participate were more likely to participate for multiple years. Nonwhite beneficiaries were 10 percentage points less likely to participate than white beneficiaries and nonwhite beneficiaries who participated for at least one year were 5.5 percentage points less likely to participate for more than one year. Dual eligibles were 7 percentage points less likely to participate. Inpatient service use was not associated with

[41] Interpretation of the coefficient for HRA risk level is not meaningful because one vendor used a 100 point scale while the other used a 17 point scale, and the scales were ordinal, not cardinal.

participation; but higher Medicare payments were associated with lower rates of participation and higher rates of multi-year participation, though the effects were very small (less than 1 percentage point per $1,000 of spending).

Vendor B's participation decreased among individuals with Alzheimer's (approximately 9 percentage points, p-value = 0.000) and increased among beneficiaries with arthritis (approximately 4 percentage points, p-value = 0.000). Higher overall risk levels as measured by HRA responses were associated with decreased likelihood of participation in multiple years.[42] Also, rating oneself as having poor health was associated with a 3 percentage point increase (p-value = 0.000). Assignment to treatment Arms 1 and 2 was negatively associated with participation in multiple years.

Table 6: Vendor B – Participation Regression Results

	Any Participation		Multi-Year Participation	
	Coefficient	P-value	Coefficient	P-value
(1)	(2)	(3)	(4)	(5)
Female	0.005	0.446	0.015	0.192
Age	0.001	0.729	0.006**	0.020
Nonwhite	-0.101***	0.000	-0.055***	0.002
Dual eligible	-0.071***	0.000	0.011	0.621
Total Medicare payment	-0.0000*	0.079	0.000**	0.019
Any inpatient claims	0.011	0.273	-0.031*	0.089
Alzheimer's	-0.092***	0.000	-0.066*	0.093
Cancer	0.005	0.632	0.011	0.582
Heart failure	-0.014	0.248	-0.023	0.322
Chronic kidney disease	-0.014	0.223	-0.013	0.563
COPD	-0.007	0.547	0.001	0.968
Depression	-0.025**	0.028	0.020	0.339
Diabetes	-0.013*	0.080	0.012	0.399
Ischemic heart disease	0.004	0.592	0.036**	0.009
Osteoporosis	0.041***	0.002	-0.009	0.695
Rheumatoid/osteoarthritis	0.038***	0.000	0.004	0.737
Stroke/TIA	-0.013	0.479	-0.059	0.100
HCC community risk score	0.002	0.756	-0.037***	0.001
HRA Risk Level			-0.003***	0.000
Poor self-rated health status			0.030*	0.060
Arm 1			-0.135***	0.000
Arm 2			-0.160***	0.000
N	25979		7335	
R-sq	0.016		0.036	

Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels

[42] Interpretation of the coefficient for HRA risk level is not meaningful because one vendor uses a 100 point scale while the other uses a 17 point scale, and the scales are ordinal, not cardinal.

Table 7 present results of regressions explaining whether beneficiaries in the refresh sample differed from those in the original sample for Vendor A. Generally, we did not find statistically significant differences between the two. The findings were similar for Vendor B (Table 8). In the relatively few cases where there were statistically significant associations, the effects were small. For example, beneficiaries with arthritis were 1 percentage point less likely to be in the original sample for Vendor A (p-value = 0.010) and Vendor B (p-value = 0.022); females were 0.67 percentage points more likely to be in the original sample for Vendor A (p-value = 0.083); beneficiaries with diabetes were 0.8 percentage points less likely to be in Vendor B's original sample (p-value = 0.069).

Table 7. Vendor A – Original versus Refresh Sample Participation Results[a]

	All Demonstration Members		Participants Only	
	Coefficient	P-value	Coefficient	P-value
Female	0.007*	0.083	0.002	0.454
Age	0.000	0.741	0.000	0.657
Nonwhite	0.005	0.322	0.004	0.218
Dual eligible	-0.002	0.757	-0.009**	0.028
Total Medicare payment	0.000***	0.004	0.000	0.396
Any inpatient claims	-0.008	0.223	-0.003	0.460
Alzheimer's	0.010	0.344	-0.016**	0.028
Cancer	0.012*	0.083	0.003	0.522
Heart failure	-0.004	0.618	0.005	0.258
Chronic kidney disease	0.003	0.727	-0.003	0.438
COPD	0.007	0.349	0.000	0.953
Depression	0.008	0.242	-0.001	0.788
Diabetes	-0.005	0.236	-0.003	0.250
Ischemic heart disease	0.004	0.454	0.004	0.104
Osteoporosis	-0.022***	0.007	-0.005	0.248
Rheumatoid/osteoarthritis	-0.010**	0.018	0.002	0.419
Stroke/TIA	0.024**	0.047	0.010	0.211
HCC community risk score	-0.002	0.602	-0.002	0.469
HRA Risk Level			0.001	0.288
Poor self-rated health status			0.004	0.207
Arm 1			-0.002	0.482
Arm 2			0.001	0.777
N	51763		7304	
R-sq	0.001		0.003	

[a] The dependent variable is an indicator for being in the original versus the refresh sample. The indicator equals 1 for being in the original sample and 0 otherwise. Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Table 8. Vendor B - Original versus Refresh Sample Participation Results[a]

	All Demonstration Members		Participants Only	
	Coefficient	P-value	Coefficient	P-value
Female	-0.004	0.252	0.001	0.267
Age	0.001	0.321	0.000	0.983
Nonwhite	0.007	0.214	0.000	0.841
Dual eligible	-0.008	0.230	0.004	0.156
Total Medicare payment	0.000***	0.001	0.000	0.455
Any inpatient claims	0.006	0.295	0.003	0.203
Alzheimer's	-0.001	0.932	0.003	0.496
Cancer	-0.001	0.919	0.000	0.849
Heart failure	-0.007	0.361	-0.001	0.683
Chronic kidney disease	0.004	0.574	0.001	0.677
COPD	0.008	0.267	0.002	0.472
Depression	0.010	0.160	-0.001	0.711
Diabetes	-0.008*	0.069	0.001	0.581
Ischemic heart disease	-0.003	0.540	0.000	0.773
Osteoporosis	0.003	0.735	-0.005*	0.062
Rheumatoid/osteoarthritis	-0.010**	0.022	0.002	0.224
Stroke/TIA	0.012	0.290	-0.009**	0.025
HCC community risk score	-0.003	0.475	0.000	0.771
HRA Risk Level			-0.000*	0.092
Poor self-rated health status			0.002	0.242
Arm 1			-0.001	0.403
Arm 2			0.001	0.762
N	51760		7335	
R-sq	0.001		0.003	

[a] The dependent variable is an indicator for being in the original versus the refresh sample. The indicator equals 1 for being in the original sample and 0 otherwise. Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels. Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

3.6 Discussion and Conclusions

Despite SRRD's randomized design, the voluntary nature of the intervention required careful analyses of a number of questions related to participation, including rates of participation across beneficiary characteristics, duration of participation, baseline characteristics of participating beneficiaries versus those of beneficiaries who chose not to participate, and baseline characteristics of individuals included in the original sample versus those of the refresh sample added at the beginning of Year 2. Answers to these questions can provide useful information and lessons learned for future voluntary interventions directed at Medicare beneficiaries. Also, information about participation rates and baseline beneficiary

characteristics provides important context for the interpretation of estimated program impacts on a range of outcomes.

Most participation-related patterns were similar across Vendors A and B. Rates of participation did not differ by gender, age, inpatient utilization, and number of chronic conditions. However, beneficiaries who were members of minority groups, those who were dually eligible, and those with total Medicare expenditures falling within the top and bottom quartiles were less likely to enroll in the demonstration. Future wellness efforts may wish to consider ways to more effectively recruit members of these groups. Individuals with cancer, osteoporosis, and arthritis were more likely to participate, while beneficiaries with the chronic conditions of Alzheimer's disease or stroke/TIA were less likely to participate. The latter result is consistent with SRRD's requirement that beneficiaries have sufficient mental capacity to participate.

Vendor B's IG had high baseline inpatient and total expenditures. This resulted in the finding that Vendor B's participants had lower expenditures than did its IG. Vendor A's IG and participants had statistically indistinguishable baseline expenditures.

In general, characteristics of beneficiaries who participated in the demonstration for at least one year were similar to those of beneficiaries participating in multiple years. The only exception was for race/ethnicity: Multi-year participants were more likely to be non-Hispanic whites. We also found that participants assigned to the "no treatment" Arm 3 had longer average duration of participation than did participants assigned to the more intensive Arms 1 or 2. The more intensive treatment arms may plausibly have had lower retention than Arm 3 because either (1) beneficiaries found participating in the intervention burdensome or (2) the benefits of the intervention accrued quickly (so that wellness goals were achieved quickly) and participants considered continued enrollment to have small additional benefit. Future wellness interventions should consider exploring strategies to ensure that the interventions remain engaging. This may include experimenting with the modalities by which coaching is delivered or taking additional steps to ensure that topics are of interest to participants. Beneficiary characteristics were generally comparable between the original versus refresh sample.

4. IMPACT OF SRRD ON CLAIMS-BASED HEALTH OUTCOMES

4.1 Introduction

In this chapter we assess the impact of SRRD on program-relevant health outcomes constructed from the Medicare administrative claims data for each vendor. The SRRD design allowed for a rigorous comparison of outcomes between randomized treatment and control groups before and after the start of the demonstration. We examined SRRD impacts for the overall demonstration and each year within the demonstration, separately by vendor and by National and Local samples. We also examined the effect of using two alternative regression specifications and explored the possible presence of placebo effects.

Summary of Results

- The impacts of interest include Medicare expenditures, Medicare use, and certain preventive screenings for Vendor A's and Vendor B's National and Local samples.
- There were no overall impacts of the demonstration for either vendor; however, upon closer examination of the yearly impacts of the demonstration, we found that:
  - For Vendor A's National sample, individuals who received the intervention starting in Year 1 of the demonstration had lower Medicare expenditures in that first year.
    - Total Medicare spending was $655 lower for the treatment group than the control group.
    - Medicare inpatient payment was $479 lower for the treatment group than the control group.
    - These impacts were not sustained in later years.
  - For Vendor B's Local sample, SRRD impacts on individuals receiving the intervention in Year 2 include the following:
    - Medicare expenditures and Medicare use were lower and preventive screening rates were higher.
    - These impacts did not continue into Year 3.
- We used two alternative estimators—difference-in-differences and a two-part model—and the results were not sensitive to the estimator used.
- There was evidence of "placebo effects" for Vendor A's Local sample—Medicare expenditures and Medicare use were lower during the demonstration period among those invited to participate, and these effects were not explained by actual participation.

4.2 Data and Methodology

4.2.1 SRRD Design

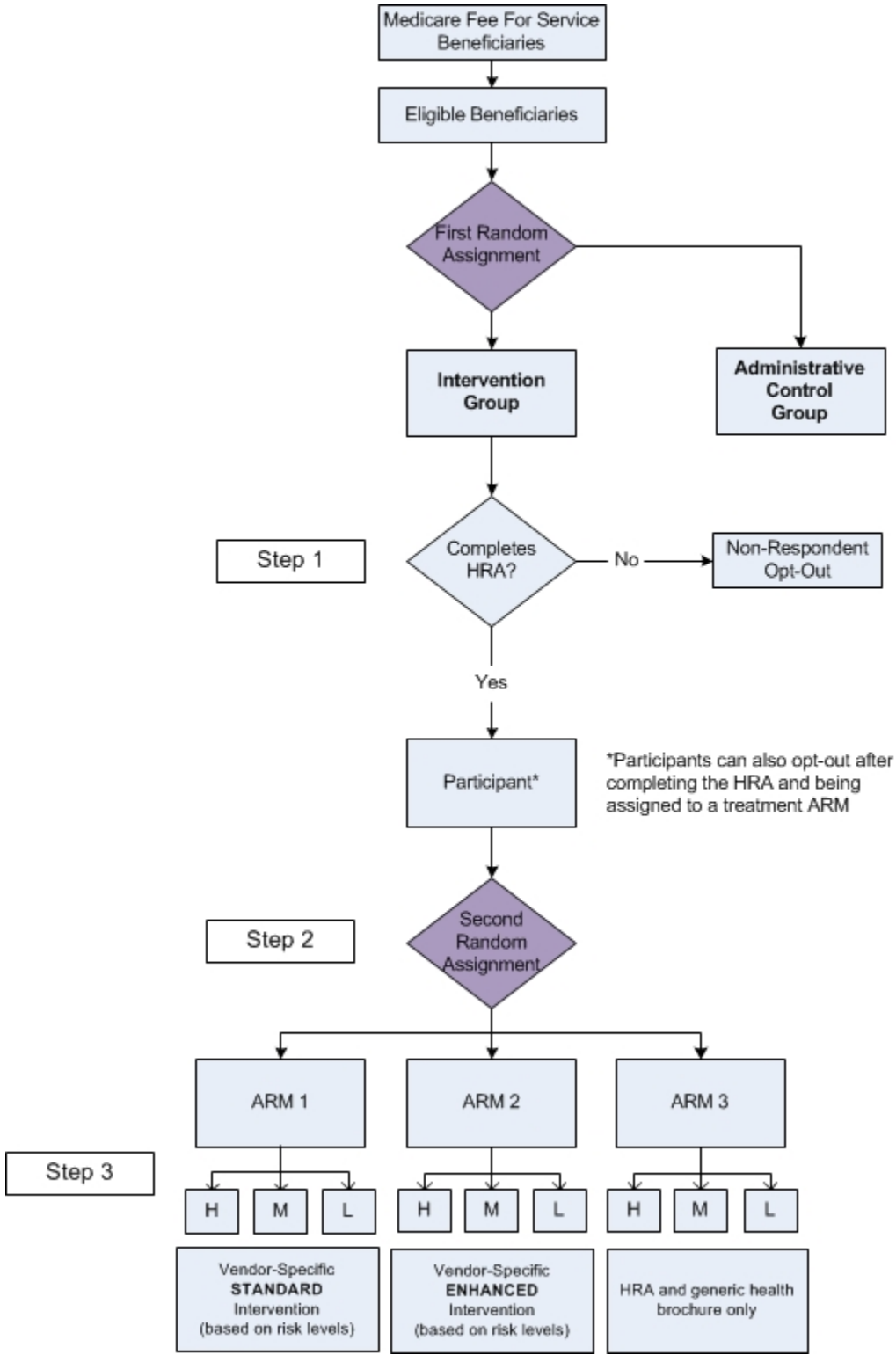
The SRRD demonstration had a RCT design with two rounds of random assignment. For each vendor, 40,000 eligible Medicare beneficiaries were randomly assigned to an Intervention Group (IG) or an Administrative Control Group (ACG) (20,000 in each group).[43] Each vendor then mailed health risk assessment (HRA) packets to its assigned IG beneficiaries to recruit them for the study (i.e., invite them to participate). Figure 1 provides an overview of the SRRD design. Once a vendor received a completed HRA from a beneficiary, the beneficiary was considered successfully recruited into the study and became an SRRD participant (Step 1 from Figure 1). Participants were then randomized a second time into one of three Intervention Arms for each vendor (Step 2 from Figure 1):

- Arm 1 – Standard Treatment (HRA + standard tailored follow-up)
- Arm 2 – Enhanced Treatment (HRA + enhanced tailored follow-up)
- Arm 3 – HRA Only (HRA + generic health advice).

Arms 1 and 2 participants received varying levels of tailored intervention services aimed at reducing risk and improving health, depending on the Arm to which they were randomly assigned and each beneficiary's HRA responses. Arm 3 participants received only a generic health mailing. The SRRD demonstration started May 1, 2009 and ended April 31, 2012.

[43] Vendors received a "refresh" sample consisting of an additional 11,800 individuals for each vendor at the end of Year 1 of the demonstration.

Figure 1. SRRD Study Design



Note: H = High Risk | M = Moderate Risk | L = Low Risk

Each vendor implemented the program nationally and in two selected local areas in partnership with the local ADRCs, which reached out to participants and provided them with additional information on ADRC services and available community resources.

The maximum amount of time during which a beneficiary could participate was two years for Vendor A (since Vendor A withdrew from the demonstration early) and three years for Vendor B. Tables 1 and 2 report the number of participants in each treatment Arm's National and Local samples for each vendor. Overall, 9,340 individuals participated in Vendor A's program and 8,717 in Vendor B's.

Table 1. Vendor A Number of Participants by Treatment Arm[a]

	Vendor A		
	National	**Local**	**Total**
Treatment Arm 1 – Standard	2,528	597	3,125
Treatment Arm 2 – Enhanced	2,525	588	3,113
Treatment Arm 3 – Mailing only	2,514	588	3,102
Total	**7,567**	**1,773**	**9,340**

[a]Participants consisted of individuals submitting one or more HRAs during the demonstration. Participants becoming ineligible during the demonstration period are excluded.

Table 2. Vendor B Number of Participants by Treatment Arm[a]

	Vendor B		
	National	**Local**	**Total**
Treatment Arm 1 – Standard	2,334	548	2,882
Treatment Arm 2 – Enhanced	2,359	559	2,918
Treatment Arm 3 – Mailing only	2,368	549	2,917
Total	**7,061**	**1,656**	**8,717**

[a]Participants consisted of individuals submitting one or more HRAs during the demonstration. Participants becoming ineligible during the demonstration period are excluded.

This evaluation used an intent-to-treat framework, in which outcomes for each vendor were compared between the randomized treatment and control groups, with additional controls for any baseline differences in beneficiary characteristics. We focused our impact analyses on participants (Step 2 of Figure 1). These individuals chose to participate in the program and were randomized into treatment (Arms 1 and 2) or control (Arm 3) groups.

4.2.2 Data Sources

The Medicare administrative data used for this evaluation includes claims from Medicare Part A and B, demographic and enrollment files, and information on chronic conditions available from the Chronic Condition Warehouse (CCW).

We also obtained additional information from CMS and vendors on beneficiaries assigned to Vendors' IGs and ACGs, as well as information on beneficiaries who agreed to participate in the demonstration. Using a linkable beneficiary identifier, we were able to combine beneficiaries' data with their Medicare administrative data.

4.2.3 Claims-based Outcome Measures

We constructed three types of claims-based outcome measures:

- Medicare payment (six measures)
- Medicare use (five measures)
- Preventive screening (three measures).

Table 3 lists the specific measures along with their data sources.

Table 3. Outcome Measures and Data Sources

Category	Measures	Data Source
Medicare payment (dollars)	Total Medicare payment Inpatient payment Outpatient payment Carrier payment Emergency department (ED) payment Office-based physician visits payment	Medicare claims from January 1, 2008 to April 30, 2012
Medicare use	Any inpatient admissions Inpatient days Any outpatient visits Office-based physician visits Any ED visit	Medicare claims from January 1, 2008 to April 30, 2012
Preventive services use	Colorectal cancer screening Breast cancer screening Cardiovascular disease screening	Medicare claims from January 1, 2008 to April 30, 2012

Medicare use and expenditure measures capture the use and cost of services, overall and by each of three settings (inpatient, outpatient, and carrier). We also included whether the

beneficiary had any emergency department (ED) visits and total Medicare payments for ED visits.

Preventive screening measures include binary indicators for beneficiaries who obtained colorectal cancer screening, breast cancer screening, and cardiovascular screening. We identified the set of preventive screenings recommended and covered by Medicare using three criteria:

- The measure could readily be evaluated using Medicare claims.
- Use of the preventive screening service was indicated for a large proportion of Medicare beneficiaries.
- Use of the preventive screening service was associated in some way, directly or indirectly, with the vendor interventions aimed at good health.

Using the Healthcare Common Procedure Coding System (HCPCS); the Current Procedural Terminology (CPT) codes; and the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes, we identified these three preventive screenings in the Medicare administrative claims data. For each preventive screening measure we excluded beneficiaries with that condition, which is consistent with screening for undiagnosed cases of the disease.[44] For example, we calculated colorectal cancer screening rates on the population of beneficiaries without a known diagnosis of colorectal cancer, as identified in the CCW chronic condition files for 2009. Beneficiaries were considered as having been screened for colorectal cancer, for example, if any one of several services was used (fecal occult blood test annually, flexible sigmoidoscopy every five years, or colonoscopy every 10 years).

The acceptable periodicity varies for each of these colorectal cancer screening modalities. Our analysis involved searching the seven Medicare claims files[45] for evidence of any of these screening modalities. Since our method did not include review of 10 years of claims data, it is possible that we have underreported rates of colorectal cancer screening for patients screened with colonoscopy 10 years earlier. However, if such underreporting were present, randomization would ensure that this bias would be distributed comparably across both treatment and control groups. Since we did not find screening rates to vary across treatment and control groups during the baseline period, we concluded that any underreporting did not introduce bias.

4.2.4 Exclusions

[44] CMS (2009). *The Guide to Medicare Preventive Services for Physicians, Providers, Suppliers, and Other Health Care Professionals*, 3rd ed.

[45] The seven Medicare claims file types are inpatient, outpatient, skilled nursing facility, hospice, home health agency, carrier, and durable medical equipment.

We excluded from the analysis individuals who enrolled in a Medicare Advantage health plan (i.e., was no longer a fee-for-service [FFS] beneficiary) or died at any time during the demonstration. We excluded beneficiaries who switched from Parts A and B to Medicare Advantage, because we could no longer observe their Medicare utilization and costs. We excluded individuals who died during the post-implementation period, because their claims histories provided upwardly skewed cost and utilization information (both because deceased individuals lacked data for the entire post-period and because health care expenditures are often concentrated at the end of life).[46]

While we wanted to eliminate the sources of bias described above, we also recognized that whether treated beneficiaries had a higher likelihood of becoming ineligible (through either death, loss of benefits, or movement to Medicare Advantage) than did control group beneficiaries is a substantively important question. In order to investigate it, we used a multinomial logit model to analyze beneficiaries' likelihood of falling into one of four categories: always eligible (the base category), ineligible due to death, ineligible due to enrollment in Medicare Advantage, or ineligible due to loss of Part A and/or Part B FFS benefits for some other reason.

We found that individuals in the IG were no more likely to become ineligible than individuals in the ACG—with similar results for participation within treatment Arms and for participation in each of the National and Local samples. While these results indicate that inclusion in the treatment group was not associated with becoming ineligible, other covariates did affect the likelihood of ineligibility. As expected, older beneficiaries and those with multiple chronic conditions were more likely to die and less likely to enter Medicare Advantage. Also, females were less likely to die; nonwhites were less likely to die and were more likely to enter Medicare Advantage or lose FFS benefits for some other reason; and dual eligibles were more likely to become ineligible due to death, enrollment in a Medicare Advantage plan, and through loss of Part and/or Part B for some other reason.

To preserve the representativeness of the original sample, we re-weighted the sample after dropping ineligible beneficiaries to match the original distribution based on these characteristics and re-ran all the analyses. Since we did not find any difference between the results using weighted regression versus non-weighted regression, all results reported below are unweighted.

[46] For example, consider an individual who dies one year after entering the demonstration. Expenditures during the year prior to death are likely to be higher than average but would be used to represent the individual's average annual expenditures during the post-period. Even though this evaluation used a randomized controlled design and even if the randomization was successful, it is possible that an imbalance in which the treatment group had more participants who died during the demonstration could skew the results toward an inappropriate finding that the demonstration increased expenditures and utilization.

4.2.5 Regression Models

Our regression-based approach enabled us to estimate the demonstration's impact (i.e., outcome differences between treatment and control groups) while controlling for baseline beneficiary characteristics. An additional benefit of this regression approach over a simple comparison of means is increased precision of the estimates, enabling detection of statistically significant results that might otherwise have been missed.

Equation 1 describes the main regression model we used. Regressions included only those individuals who completed at least one HRA. Once a beneficiary was a participant (completed a HRA), we counted him/her as a participant for the duration of the demonstration. The regression was run separately for each vendor and included the following control variables:

- Demographic characteristics
  - Age
  - Female (indicator variable)
  - Nonwhite (indicator variable)
- Medicare-Medicaid dual eligibility (indicator variable)
- Two or more chronic conditions (indicator variable)
- Baseline (lagged) value of the dependent variable
- Total Medicare payments from the baseline year:

Equation 1.

$$y_{i,t=2} = \delta_0 + \delta_1 y_{i,t=1} + \delta_2 totalpayment_{i,t=1} + \delta_3 age_{i,t=1} + \delta_4 female_i + \delta_5 nonwhite_i + \delta_6 dual_{i,t=1} + \delta_7 multi\_cc_i + \delta_8 Treatment\_Arm_i$$

where:

$Y_{i,t=2}$ = the average yearly outcome of interest for participant *i* during the demonstration period
$Y_{i,t=1}$ = the outcome of interest in the base period
$Totalpayment_{i,t=1}$ = total Medicare expenditures in the base period
$age_{i,t=1}$ = beneficiary's age in the base period
$female_i$ = indicator for female versus male
$nonwhite_i$ = indicator for nonwhite race/ethnicity
$dual_{i,t=1}$ = indicator for dual eligibility
$multi\_cc_{i,t=1}$ = indicator for two or more chronic conditions
$Treatment\_Arm_i$ = indicator for receiving the intervention (i.e. randomization into either Arm 1 or Arm 2).

We estimated the parameters using ordinary least squares, which models the outcome as a linear function of the predictors. The coefficient of interest is the one on the *Treatment_Arm* indicator, which provides the impact of being in either Arm 1 or Arm 2 compared with the

omitted category (Arm 3) and is the estimated overall impact of the demonstration. The main results focus on Arm 1 and Arm 2 combined, since the purpose of the evaluation is to examine the impact of risk reduction programs with follow-up interventions, and pooling Arms provides additional power for the detection of impacts. We did also examine impacts separately for Arm 1 and Arm 2 to detect any differential impacts of the more intense Arm 2 intervention. Since we found no additional impacts of Arm 2 participation over Arm 1, the results reported in this chapter focus on pooled Arm 1 and 2 results.

Yearly Regressions. Since the main purpose of this study is to estimate the overall impacts of the demonstration, we combined all demonstration years into one "demonstration period" and estimated the average yearly impact over the entire demonstration period. However, we also examined how impacts varied by year to gain insight into the timing and pattern of impacts. We ran separate regressions for each of the three demonstration years for Vendor A and Vendor B.[47]

First Alternative Method: Difference-in-Differences. We implemented a difference-in-differences (DID) methodology. Since beneficiaries were first randomized into the IG and ACG, and those who chose to participate were then randomized into Arms 1, 2, and 3, regression adjustment for beneficiary characteristics is not essential to obtain unbiased impact estimates. The DID for outcome Y is given in the following equation (where t=2 refers to the intervention [post] period, t=1 refers to the baseline [pre] period, and $\bar{Y}$ denotes the mean of Y):

Equation 2. $$Y = \left(\bar{Y}_{t=2,Arm12} - \bar{Y}_{t=2,Arm3}\right) - \left(\bar{Y}_{t=1,Arm12} - \bar{Y}_{t=1,Arm3}\right)$$

The expression in the first set of parentheses is the difference in the mean of *Y* during the intervention period for pooled Arms 1 and 2 participants versus Arm 3 participants. From this, the DID estimate subtracts the second expression, which is the pooled Arms 1 and 2 minus Arm 3 difference in *Y* during the baseline period. Subtracting the pooled Arms 1 and 2 minus Arm 3 difference during the baseline period from that during the intervention period removes any time-invariant differences in outcomes between the treatment (pooled Arms 1 and 2) and control (Arm 3) groups that would exist regardless of time period (pre versus post). This is important in cases where there could be differences in the treatment and control groups that do not change over time (e.g., genetic predisposition to illness).

Second Alternative Method: Two-Part Model

A large body of literature addresses the skewed nature of medical expenditures since, depending on the length of the observation window, many people will have zero medical expenditures. A large mass at zero skews the distribution of expenditures and can lead to potentially biased and less precise results (especially in smaller samples). Although zero total

[47] Although Vendor A left the demonstration after Year 2, the data do allow for the estimation of impacts in Year 3 on the previous participants, and we did explore this.

expenditure is less of an issue among the Medicare population than among lower-expenditure populations, it is still an issue for rarer outcomes such as inpatient payments.

One common method used to address skewed medical expenditures is the two-part model (2PM) in which the first part addresses whether or not the individual received any medical care (the probability of getting any care) and the second part addresses the amount of expenditures on only those with some amount of care. This model can be written mathematically as:

Equation 3. $E(Y|x) = \Pr(Y > 0) \times E(Y|Y > 0)$

The first stage, $\Pr(Y > 0)$, can be modeled with a logistic regression for receiving any medical care. [48,49]

Equation 4. $Pr(Y > 0) = \frac{e^{x\prime\beta}}{1+e^{x\prime\beta}}$

The second stage $E(Y|Y > 0)$, which is conditional on have some medical expenditures, can then be estimated using linear regression. We estimated the two-step model for relevant payment outcomes, using the right-hand side covariates listed above for both stages of the regression.

Placebo Effects. Individuals in the IG were sent invitations to participate and those among the IG who chose to participate were then randomized into three different treatment arms, as noted, where two arms received varying levels of treatment and one arm received only a generic mailing on health (see Figure 1). The rest of the IG were non-responders (or had opted out of the demonstration). We expect the impacts of SRRD to be driven by those receiving the full risk reduction programs (i.e., participants in Arms 1 or 2). However, if receiving the HRA influences health behavior, there may be a "placebo effect" in which members of the IG who simply received (or completed but did not return) the HRA, changed their health behavior.

The SRRD design allowed us to examine any placebo effect because the ACG was an administrative control group and individuals randomized into this group were not contacted in any way. If we presume that the only source of SRRD impact is participation in Arm 1 and Arm 2, we will expect any differences in outcomes between the IG and ACG to be proportional to the number of participants in Arm 1 and Arm 2. For example, suppose that we find that Arm 2 reduces total Medicare payments by $600 per participant and Arm 2 reduces it by $400 per participant compared with Arm 3.[50] If participation is 45 percent (15 percent in each treatment arm) and Arm 1 and Arm 2 are the only source of impact on total Medicare payments, we will expect mean total Medicare payments for the IG to be lower than that of the ACG by $150

[48] http://www.unc.edu/~enorton/DebManningNortonPresentation.pdf

[49] Buntin, Melinda Beeuwkes and Alan M. Zaslavsky, "Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures," *Journal of Health Economics*, 23 (2004), 525-542.

[50] This placebo effect analysis and example is presented by the CMS contractors who designed the SRRD: Goetzel, Ron Z., Stapleton, David, Shechter, David, Livermore, Gina, Ozminkoswski, Ronald, "Senior Risk Reduction, Final Project Report," Thomson Medstat, March 2, 2004.

[(0.15 X 600) + (0.15 X 400)]. If the difference between IG and ACG mean payments is statistically significantly different from $150, it is reasonable to conclude that a placebo effect exists.

To examine the existence of such an effect, we estimated Equation 5 for each of the outcomes of interest for all demonstration beneficiaries for each vendor, using the same right-hand side control variables as in Equation 1 and adding an indicator for being in the IG:

Equation 5.
$$y_{i,t=2} = \beta_0 + \beta_1 y_{i,t=1} + \beta_2 totalpayment_{i,t=1} + \beta_3 age_{i,t=1} + \beta_4 female_i + \beta_5 nonwhite_i + \beta_6 dual_{i,t=1} + \beta_7 multi\_cc_i + \beta_8 IG_i$$

where:

$Y_{i,t=2}$ = the average yearly outcome of interest for participant *i*
$Y_{i,t=1}$ = the outcome of interest in the base period
$Totalpayment_{i,t-1}$ = total Medicare expenditures in the base period
$age_{i,t=1}$ = beneficiary's age in the base period
$female_i$ = an indicator for female versus male
$nonwhite_i$ = an indicator for being nonwhite
$dual_{i,t=1}$ = an indicator for dual eligibility
$multi\_cc_{i,t=1}$ = an indicator for two or more chronic conditions
IG_i = an indicator for IG members.

The estimated coefficient on the indicator IG_i, β_8, is the impact of being in the IG versus the ACG. To examine whether placebo effects existed we compared the estimated β_8 with the estimated effect of being in the pooled Arm 1 and Arm 2 group, weighted by the number of participants in each arm. Statistically significant differences indicate the existence of placebo effects.

4.3 Results

In this section we first report differences in means and t-tests between the pooled Arm 1 and 2 participants (HRA plus tailored programs) and Arm 3 (HRA-only) participants. We then present the results of the main regression analyses, yearly regressions, and alternative specifications. We present results separately by vendor and by National versus Local samples.

4.3.1 Statistical Tests of Means

Tables 4 through 7 present the means for the pooled Arms 1 and 2 participants and Arm 3 participants for Vendor A's and Vendor B's National and Local samples. The tables also show the differences in means during the demonstration period and the p-value from a *t*-test of whether the difference was statistically different from zero.

For Vendor A's National and Local samples (Tables 4 and 5, respectively), there were no statistically significant differences between any of the means in the demonstration period. Participants in the Arms 1 and 2 incurred average yearly Medicare expenditures during the demonstration period of $6,536, and participants in Arm 3 incurred average yearly Medicare expenditures during the demonstration period of $6,361. The difference of $175 was not statistically significant (p-value = 0.562). Further, there were no clear patterns in the signs of the differences for each of the outcomes, though directions were generally similar across the National and Local samples. Thus, examining only demonstration period means shows no SRRD impact (i.e., no differences in outcomes between the treatment and control groups).

Table 4. Vendor A National Outcomes Means and T-Tests by Treatment Arm[a]

	Pooled Arms 1 & 2	Arm 3	Arms 1 & 2 - Arm 3 Difference	P-Value
Sample Size	5,053	2,514	-	-
Payment Variables				
Total Medicare payment	$6,536	$6,361	$175	0.562
Inpatient payment	$2,156	$2,190	$-34	0.856
Outpatient payment	$1,199	$1,136	$63	0.399
Carrier payment	$2,412	$2,389	$23	0.785
ED payment	$96	$88	$7	0.215
Office visit payments	$1,438	$1,483	$-45	0.495
Utilization Variables				
Any inpatient visits	21.9%	21.7%	0.17%	0.866
Inpatient days	0.91	0.94	-0.02	0.783
Any outpatient visits	81.6%	80.9%	0.67%	0.484
Office days	13.80	13.63	0.17	0.562
Any ED visits	33.8%	32.1%	1.62%	0.157
Preventive Screening				
Colorectal cancer	31.2%	30.1%	1.06%	0.346
Breast cancer	68.6%	70.0%	-1.42%	0.358
Cardiovascular disease	79.5%	78.4%	1.08%	0.359

[a] The demonstration period was May 1, 2009 to April 30, 2011 and the baseline period was January 1, 2008 to April 30, 2009. Variable values have been annualized.

Table 5. Vendor A Local Outcomes Means and T-tests by Treatment Arm[a]

	Pooled Arms 1 & 2	Arm 3	Arm 1 & 2 - Arm 3 Difference	P-Value
Sample Size	1,185	588	-	-
Payment Variables				
Total Medicare payment	$5,946	$5,819	$126	0.826
Inpatient payment	$1,907	$2,054	$-147	0.640
Outpatient payment	$1,358	$1,209	$149	0.325
Carrier payment	$1,966	$1,906	$60	0.593
ED payment	$95	$84	$11	0.274
Office visit payments	$1,097	$1,074	$23	0.732
Utilization Variables				
Any inpatient visits	20.2%	19.4%	0.78%	0.697
Inpatient days	0.71	0.72	-0.02	0.885
Any outpatient visits	87.9%	84.7%	3.24%	0.066
Office days	12.34	11.90	0.44	0.456
Any ED visits	33.4%	30.1%	3.32%	0.156
Preventive Screening				
Colorectal cancer	31.1%	31.0%	0.08%	0.974
Breast cancer	74.2%	77.0%	-2.79%	0.348
Cardiovascular disease	82.9%	79.5%	3.35%	0.143

[a] The demonstration period was May 1, 2009 to April 30, 2011 and the baseline period was January 1, 2008 to April 30, 2009. Variable values have been annualized.

Similarly, for Vendor B, there were no statistically significant differences between means in the outcome variables between the pooled Arms 1 and 2 participants and Arm 3 participants in the National sample, although all mean Medicare payment variables were lower for pooled Arms 1 and 2 participants than for Arm 3 participants. Pooled Arms 1 and 2 participants in the National sample incurred $64 (p-value = 0.833) less than Arm 3 participants in Medicare expenditures (Table 6). For Vendor B's local sample, this difference is even larger: pooled Arms 1 and 2 participants spent $1,300 (p-value = 0.048) less than Arm 3 participants (Table 7) and the difference was statistically significant at the 5 percent significance level. Outpatient payments were $565 (p-value = 0.015) lower for pooled Arms 1 and 2 participants than for Arm 3 participants. There were also positive differences in screening rates. Pooled Arms 1 and 2 participants were 4.70 percentage points (p-value = 0.057) more likely to be screened for cardiovascular disease than Arm 3 participants.

Caution is indicated in interpreting these results, however, because there were differences (though generally not statistically significant) in baseline values for Vendor B in its Local sample. In Vendor B's Local sample, pooled Arms 1 and 2 participants spent $867 less (p-value = 0.217) than Arm 3 participants in the baseline period. These differences in baseline values suggest the importance of controlling for baseline characteristic in the impact analyses.

Table 6. Vendor B Outcomes Means and T-tests by Treatment Arm [a]

	Pooled Arms 1 & 2	Arm 3	Arm 1 & 2 - Arm 3 Difference	P-Value
Sample Size	4,693	2,368	-	-
Payment Variables				
Total Medicare payment	$6,776	$6,840	$-64	0.833
Inpatient payment	$2,268	$2,326	$-59	0.729
Outpatient payment	$1,258	$1,271	$-13	0.862
Carrier payment	$2,480	$2,463	$18	0.845
ED payment	$88	$92	$-4	0.398
Office visit payments	$1,518	$1,487	$32	0.655
Utilization Variables				
Any inpatient visits	28.4%	29.4%	-1.05%	0.358
Inpatient days	0.91	1.01	-0.09	0.289
Any outpatient visits	87.8%	87.0%	0.84%	0.318
Office days	13.85	13.69	0.17	0.597
Any ED visits	41.9%	43.0%	-1.06%	0.397
Preventive Screening				
Colorectal cancer	40.4%	40.1%	0.32%	0.796
Breast cancer	75.9%	73.9%	2.07%	0.177
Cardiovascular disease	85.3%	86.1%	-0.86%	0.415

[a] The demonstration period was May 1, 2009 to April 30, 2012 and the baseline period was January 1, 2008 to April 30, 2009. Variable values have been annualized.

Table 7. Vendor B Local Outcomes Means and T-tests by Treatment Arm[a]

	Pooled Arms 1 & 2	Arm 3	Arm 1 & 2 - Arm 3 Difference[b]	P-Value
Sample Size	1,107	549	-	-
Payment Variables				
Total Medicare payment	$6,716	$8,017	$-1,300**	0.048
Inpatient payment	$2,203	$2,602	$-400	0.217
Outpatient payment	$1,139	$1,704	$-565**	0.015
Carrier payment	$2,668	$2,910	$-243	0.271
ED payment	$78	$87	$-9	0.332
Office visit payments	$1,099	$1,212	$-113	0.312
Utilization Variables				
Any inpatient visits	31.0%	33.3%	-2.35%	0.337
Inpatient days	0.98	1.56	-0.58	0.244
Any outpatient visits	88.4%	88.5%	-0.09%	0.958
Office days	9.83	10.05	-0.22	0.625
Any ED visits	39.4%	39.5%	-0.14%	0.956
Preventive Screening				
Colorectal cancer	29.8%	28.0%	1.78%	0.453
Breast cancer	73.9%	69.9%	4.07%	0.220
Cardiovascular disease	84.1%	79.4%	4.70%*	0.057

[a] The demonstration period is May 1, 2009 to April 30, 2012 and the baseline period is January 1, 2008 to April 30, 2009. Variable values have been annualized.
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

4.3.2 Regression results

Tables 8 through 11 present the results of regressions using Equation 1 for estimating the impact of participation in treatment Arms 1 or 2 compared with participation in the "HRA only" Arm 3. For each of the tables, the estimated coefficients, shown in column 1, represent the effect of being in either Arm 1 or Arm 2 (pooled) on the outcome of interest versus being in Arm 3. Column 2 reports the p-values associated with each estimated coefficient. Columns 3 and 4 show the regression adjusted means for the pooled Arms 1 and 2 participants and Arm 3 participants. Column 5 provides the estimated coefficient as a percentage of the Arm 3 regression adjusted mean (i.e., the percentage impact).

Vendor A. As shown in Table 8, there was a statistically significant impact of being in Arms 1 or 2 versus being in Arm 3 for only one outcome variable. Participating in Arms 1 or 2 was associated with lower probabilities of breast cancer screening compared with Arm 3 participants (-2.6 percentage points) in Vendor A's National sample. However, given the lack of

statistical significance for other variables, the varying directions on the coefficients' signs, and the likelihood that some coefficients will be significant due to chance, we do not interpret this finding as evidence of the demonstration's impact. We did not find any impacts for Vendor A's Local sample either (as indicated in Table 9).

Table 8. Vendor A National Results, Treatment Arms[a]

	Coefficient on Pooled Arms 1 and 2[b]	P-value	Reg Adj Mean, Pooled Arms 1 and 2	Reg Adj Mean Arm, Arm 3	Percent Impact
	(1)	(2)	(3)	(4)	(5)=100x(1)/(4)
Medicare Payment					
Total Medicare Payment	136	0.640	6,531	6,396	2.1%
Inpatient Payment	-8	0.967	2,145	2,153	-0.4%
Outpatient Payment	48	0.468	1,206	1,158	4.1%
Carrier payment	44	0.569	2,430	2,386	1.8%
ED payment	6	0.247	96	90	6.6%
Office visit payments	-6	0.920	1,460	1,466	-0.4%
Medicare Use					
Any inpatient visits	-0.1%	0.875	20.3%	20.4%	-0.7%
Inpatient days	-0.0	0.816	0.9	0.9	-2.2%
Any outpatient visits	0.3%	0.733	80.1%	79.8%	0.4%
Office days	0.0	0.941	13.8	13.7	0.1%
Any ED visits	1.3%	0.233	32.0%	30.7%	4.2%
Preventive Screening					
Colorectal cancer	1.1%	0.328	29.4%	28.3%	3.8%
Breast cancer	-2.6%*	0.053	65.8%	68.4%	-3.8%
Cardiovascular disease	1.2%	0.315	54.1%	52.9%	2.2%

[a] The sample for the regression consists of all Vendor A participants (individuals who completed at least one HRA). Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Table 9. Vendor A Local Results, Treatment Arms[a]

	Coefficient on Pooled Arms 1 & 2[b]	P-value	Reg Adj Mean, Pooled Arms 1 & 2	Reg Adj Mean Arm, Arm 3	Percent Impact
	(1)	(2)	(3)	(4)	(5)=100x(1)/(4)
Medicare Payment					
Total Medicare Payment	103	0.860	6,011	5,908	1.7%
Inpatient Payment	-195	0.562	1,927	2,122	-9.2%
Outpatient Payment	197	0.178	1,398	1,201	16.4%
Carrier payment	10	0.922	1,954	1,944	0.5%
ED payment	12	0.290	98	86	13.4%
Office visit payments	-1	0.979	1,088	1,089	-0.1%
Medicare Use					
Any inpatient visits	0.8%	0.667	19.0%	18.2%	4.5%
Inpatient days	-0.1	0.659	0.7	0.8	-8.1%
Any outpatient visits	2.3%	0.136	86.7%	84.4%	2.7%
Office days	0.0	0.998	12.2	12.2	0.0%
Any ED visits	3.4%	0.124	31.8%	28.4%	11.9%
Preventive Screening					
Colorectal cancer	0.7%	0.756	30.0%	29.3%	2.4%
Breast cancer	-1.8%	0.455	72.7%	74.5%	-2.4%
Cardiovascular disease	-1.6%	0.469	37.5%	39.0%	-4.0%

[a] The sample for the regression consists of all Vendor A participants (individuals who completed at least one HRA). Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Vendor B. Tables 10 and 11 show the impacts of participation in Arms 1 or 2 versus Arm 3 for Vendor B's National and Local samples, respectively. We did not find any impacts for Vendor B's pooled Arms 1 and 2 participants compared with Arm 3 participants in the National sample.

For Vendor B's Local sample (Table 11), all Medicare payment and Medicare use variables had negative coefficients. For the preventive screening outcomes, participation in Arms 1 and 2 were positively associated with higher rates of screening compared with Arm 3. However, none of the coefficients was statistically significantly different from zero, except for Medicare outpatient payments; pooled Arms 1 and 2 participants incurred $397 less in Medicare costs for outpatient visits than did Arm 3 participants. Since we would expect one or two coefficients to be statistically significant by chance given the number of outcomes we examined, we do not interpret the statistical significance of the coefficient on outpatient visits as evidence of SRRD impacts for Vendor B.

Table 10. Vendor B National Results, Treatment Arms[a]

	Coefficient on Pooled Arms 1 & 2[b]	P-value	Reg Adj Mean, Pooled Arms 1 & 2	Reg Adj Mean Arm, Arm 3	Percent Impact
	(1)	(2)	(3)	(4)	(5)=100x(1)/(4)
Medicare Payment					
Total Medicare Payment	-45	0.875	6,783	6,827	-0.7%
Inpatient Payment	-55	0.739	2,269	2,324	-2.4%
Outpatient Payment	-32	0.640	1,252	1,284	-2.5%
Carrier payment	-14	0.861	2,470	2,484	-0.6%
ED payment	-5	0.284	88	93	-5.1%
Office visit payments	3	0.966	1,509	1,506	0.2%
Medicare Use					
Any inpatient visits	-0.9%	0.422	28.4%	29.3%	-3.0%
Inpatient days	-0.1	0.274	0.9	1.0	-9.2%
Any outpatient visits	1.0%	0.182	87.9%	86.9%	1.2%
Office days	0.2	0.361	13.9	13.7	1.5%
Any ED visits	-0.9%	0.439	42.0%	42.9%	-2.2%
Preventive Screening					
Colorectal cancer	0.0%	0.974	38.7%	38.7%	0.1%
Breast cancer	1.3%	0.295	74.7%	73.3%	1.8%
Cardiovascular disease	-1.5%	0.206	59.1%	60.6%	-2.5%

[a] The sample for the regression consists of all Vendor B participants (individuals who completed at least one HRA). Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Table 11. Vendor B Local Results, Treatment Arms[a]

	Coefficient on Pooled Arms 1 & 2[b]	P-value	Reg Adj Mean, Pooled Arms 1 & 2	Reg Adj Mean Arm, Arm 3	Percent Impact
	(1)	(2)	(3)	(4)	(5)=100x(1)/(4)
Medicare Payment					
Total Medicare Payment	-958	0.111	6,816	7,774	-12.3%
Inpatient Payment	-271	0.387	2,239	2,510	-10.8%
Outpatient Payment	-397*	0.065	1,198	1,595	-24.9%
Carrier payment	-213	0.256	2,662	2,875	-7.4%
ED payment	-6	0.468	78	84	-7.4%
Office visit payments	-93	0.344	1,119	1,211	-7.6%
Medicare Use					
Any inpatient visits	-1.6%	0.477	29.0%	30.6%	-5.3%
Inpatient days	-0.6	0.266	1.0	1.5	-37.3%
Any outpatient visits	-1.1%	0.510	85.3%	86.4%	-1.3%
Office days	-0.4	0.236	9.8	10.2	-3.8%
Any ED visits	-0.2%	0.928	36.5%	36.7%	-0.6%
Preventive Screening					
Colorectal cancer	1.0%	0.664	28.1%	27.1%	3.7%
Breast cancer	1.8%	0.548	71.5%	69.7%	2.6%
Cardiovascular disease	2.5%	0.341	56.4%	53.9%	4.6%

[a] The sample for the regression consists of all Vendor B participants (individuals who completed at least one HRA). Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

4.3.3 Yearly Impacts

The results shown above combine the demonstration years into the "demonstration period." In this section, we break down the impacts by year to discern any patterns in the timing of impacts.

Vendor A. For Vendor A, examination by year reveals substantial variation among demonstration years. Although we did not find any impacts for the overall demonstration period, we did see impacts during the Year 1 for pooled Arms 1 and 2 participants versus Arm 3 participants (see Table 12). However, the impacts observed in the first year were not sustained into the second or third year (while Vendor A did not participate in the third year, Medicare claims data for the third year was available). For example, in Year 1, pooled Arms 1 and 2 participants spent $655 less than Arm 3 participants in total Medicare payments. In Year 2 and Year 3 pooled Arms 1 and 2 participants' spending was not statistically different from Arm 3 participants' spending. This pattern was similar for the other payment outcomes, as well as for the Medicare use variables. In addition, Arms 1 and 2 participants were 3.7 percentage points

less likely to have breast cancer screenings during Year 1. There were no differences in screening for Year 2 and Year 3.

Table 12. Vendor A National Yearly Results, Treatment Arms[a]

	Year 1 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 2 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 3 Coefficient on Pooled Arms 1 & 2[b]	P-value
Medicare Payment						
Total Medicare Payment	-655*	0.081	347	0.325	-99	0.827
Inpatient Payment	-479*	0.062	104	0.655	137	0.638
Outpatient Payment	-51	0.520	106	0.159	-103	0.312
Carrier payment	-174*	0.087	136	0.114	16	0.896
ED payment	1	0.852	8	0.230	1	0.849
Office visit payments	-121	0.135	57	0.385	5	0.960
Medicare Use						
Any inpatient visits	-1.1%	0.265	0.3%	0.749	1.0%	0.254
Inpatient days	-0.2**	0.040	0.0	0.783	0.1	0.469
Any outpatient visits	-0.7%	0.542	0.8%	0.426	0.1%	0.896
Office days	-0.2	0.515	-0.0	0.972	-0.2	0.395
Any ED visits	-0.6%	0.611	2.3%**	0.021	0.7%	0.509
Preventive Screening						
Colorectal cancer	1.1%	0.328	0.6%	0.512	-0.3%	0.722
Breast cancer	-3.7%**	0.035	-2.3%	0.119	-0.2%	0.913
Cardiovascular disease	-0.0%	0.978	0.9%	0.460	-0.0%	0.973

[a] For each year, the sample for the regression consists of individuals who first participated in that year or a previous year. Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Vendor B. For Vendor B, any impacts of the demonstration on pooled Arms 1 and 2 participants versus Arm 3 participants occurred in Year 2 of the demonstration (Table 13). In Year 2, pooled Arms 1 and 2 participants were 1.5 percentage points less likely to have any hospital admissions and 0.2 fewer hospital inpatient hospital days Medicare, both statistically significant at the 10 percent level.

We found substantial impacts of the demonstration for Vendor B's Local sample for Year 2. All Medicare payment outcomes were substantially lower in Year 2 for pooled Arms 1 and 2 participants versus Arm 3 participants. Total Medicare payment was $2,141 lower for pooled

Arms 1 and 2 participants than for Arm 3 participants, statistically significant at the 1 percent level. Some of the impacts appear to have continued into Year 3 as well.[51]

Table 13. Vendor B National Yearly Results, Treatment Arms[a]

	Year 1 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 2 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 3 Coefficient on Pooled Arms 1 & 2[b]	P-value
Medicare Payment						
Total Medicare payment	433	0.192	-471	0.240	104	0.799
Inpatient payment	202	0.328	-434	0.117	138	0.570
Outpatient payment	41	0.570	-50	0.522	-41	0.710
Carrier payment	148	0.157	-102	0.286	-1	0.990
ED payment	-1	0.790	-7	0.226	-5	0.500
Office visit payments	103	0.232	-20	0.781	-13	0.896
Medicare Use						
Any inpatient visits	-1.2%	0.206	-1.5%*	0.085	1.0%	0.258
Inpatient days	0.0	0.724	-0.2*	0.073	-0.0	0.714
Any outpatient visits	0.8%	0.484	0.3%	0.762	1.5%	0.141
Office days	0.2	0.359	0.0	0.918	0.4	0.148
Any ED visits	-0.5%	0.647	-1.5%	0.138	-1.1%	0.304
Preventive Screening						
Colorectal cancer	0.0%	0.974	0.3%	0.787	-0.0%	0.999
Breast cancer	2.2%	0.222	0.3%	0.847	3.5%**	0.029
Cardiovascular disease	0.1%	0.938	-0.6%	0.643	0.2%	0.897

[a] For each year, the sample for the regression consists of individuals who first participated in that year or a previous year. Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1).
[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

[51] Vendor B local sample for Year 2 regressions comprised 1,656 individuals.

Table 14. Vendor B Local Yearly Results, Treatment Arms[a]

	Year 1 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 2 Coefficient on Pooled Arms 1 & 2[b]	P-value	Year 3 Coefficient on Pooled Arms 1 & 2[b]	P-value
Medicare Payment						
Total Medicare payment	340	0.652	-2,141***	0.008	-979	0.253
Inpatient payment	-113	0.797	-730*	0.089	-38	0.944
Outpatient payment	232	0.415	-647**	0.028	-470**	0.026
Carrier payment	-27	0.896	-566**	0.046	-203	0.333
ED payment	-3	0.812	-22*	0.053	0	0.991
Office visit payments	28	0.781	-202**	0.041	-134	0.408
Medicare Use						
Any inpatient visits	-3.3%	0.122	-5.2%***	0.008	0.7%	0.678
Inpatient days	-0.2	0.422	-1.8	0.223	0.1	0.693
Any outpatient visits	-0.3%	0.915	-4.6%**	0.028	-1.2%	0.553
Office days	-0.4	0.363	-0.4*	0.071	-0.6	0.295
Any ED visits	-2.1%	0.333	-3.2%	0.125	3.1%	0.136
Preventive Screening						
Colorectal cancer	-1.1%	0.583	-2.2%*	0.091	3.0%	0.120
Breast cancer	1.5%	0.703	7.9%**	0.019	-0.3%	0.937
Cardiovascular disease	-1.5%	0.633	1.7%	0.570	4.0%	0.147

[a] For each year, the sample for the regression consists of individuals who first participated in that year or a previous year. Right-hand side variables include age, indicators for female, dual eligibility status, nonwhite, and multiple chronic conditions; as well as for participation in Arms 1 or 2 (see Equation 1)..

[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

4.3.4 Difference-in-Difference (DID) and Two Part Models (2PM)

Tables 15 and 16 compare the findings from the alternative model specifications for each vendor's National samples.

For Vendor A, the DID results (Equation 2) were generally similar to the main regression (Equation 1) results in that most coefficients were not statistically different from zero. The exception was cardiovascular disease screening. The main regression yielded a 1.2 percentage point increase that was not statistically different from zero between pooled Arms 1 and 2 participants compared with Arm 3; for the DID analyses, the difference was a 2.2 percentage point increase, statistically significant at the 10 percent level. The DID results show that breast cancer screening was more negatively associated with participation in pooled Arms 1 and 2 than the main regression indicated. The main regression found that pooled Arms 1 and 2 participants were 2.6 percentage points less like to be screened than Arm 3 participants (statistically significant at the 10 percent level). The DID results showed that pooled Arms 1 and

2 eligible participants were 4.1 percentage points less likely to be screened than Arm 3 participants (statistically significant at the 1 percent level)

For the 2PM formulation, used for certain Medicare payment variables where the mass at zero posed a risk for introducing bias into our main regression reports, the impact of participation in Arms 1 or 2 compared with Arm 3 was similar to the main regression results for Vendor A.

Table 16 reports results for Vendor B's National sample. For the Medicare payment outcome variables, the DID results did differ from the main regression results. This is likely due to larger baseline differences among Vendor B's sample. However, we did not find statistically significant impacts using the DID or the 2PM for Vendor B.

Table 15. Vendor A National, Comparison of Alternative Estimators[a]

	Main Regression	P-value	Difference-in-Difference	P-value	Two-Part Model	P-value
Medicare Payment						
Total Medicare payment	136	0.640	31	0.927	-	-
Inpatient payment	-8	0.967	-91	0.698	-14	0.938
Outpatient payment	48	0.468	46	0.509	40	0.558
Carrier payment	44	0.569	49	0.567	45	0.551
ED payment	6	0.247	3	0.580	7	0.162
Office visit payments	-6	0.920	37	0.597	-	-
Medicare Use						
Any inpatient visits	-0.1%	0.875	-0.8%	0.503	-	-
Inpatient days	-0.0	0.816	-0.1	0.520	-	-
Any outpatient visits	0.3%	0.733	0.3%	0.803	-	-
Office days	0.0	0.941	-0.0	0.877	-	-
Any ED visits	1.3%	0.233	0.2%	0.872	-	-
Preventive Screening						
Colorectal cancer	1.1%	0.328	0.6%	0.696	-	-
Breast cancer	-2.6%*	0.053	-4.1%***	0.009	-	-
Cardiovascular disease	1.2%	0.315	2.2%*	0.084	-	-

[a]Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Table 16. Vendor B National, Comparison of Alternative Estimators[a]

	Main Regression	P-value	Difference-in-Difference	P-value	Two-Part Model	P-value
Medicare Payment						
Total Medicare payment	-45	0.875	-92	0.778	-	-
Inpatient payment	-55	0.739	30	0.888	-70	0.656
Outpatient payment	-32	0.640	-54	0.468	-36	0.612
Carrier payment	-14	0.861	-60	0.521	-14	0.863
ED payment	-5	0.284	-8	0.124	-4	0.293
Office visit payments	3	0.966	-41	0.593	-	-
Medicare Use						
Any inpatient visits	-0.9%	0.422	-0.9%	0.505	-	-
Inpatient days	-0.1	0.274	-0.1	0.163	-	-
Any outpatient visits	1.0%	0.182	1.4%	0.211	-	-
Office days	0.2	0.361	0.2	0.392	-	-
Any ED visits	-0.9%	0.439	-1.1%	0.429	-	-
Preventive Screening						
Colorectal cancer	0.0%	0.974	-0.4%	0.776	-	-
Breast cancer	1.3%	0.295	1.1%	0.452	-	-
Cardiovascular disease	-1.5%	0.206	-2.4%*	0.068	-	-

[a]Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

4.3.5 Target Group Impacts—"Placebo Effects"

To examine the potential existence of a placebo effect, we estimated the impact of randomization into the IG versus the ACG using Equation 5 for each of the outcomes.

Table 17 compares the impacts of participation in pooled Arms 1 and 2 versus Arm 3 in the first two columns (duplicated from Table 9) and the effect of being in the IG versus the ACG in the third and fourth column for Vendor A's Local sample. As shown earlier there were no statistically significant differences in outcomes between pooled Arms 1 and 2 participation and Arm 3 participation. However, those in the IG spent $756 less (statistically significant at the 1 percent level) than those in the ACG in total Medicare expenditures and $478 less (statistically significant the 1 percent level) in Medicare inpatient expenditures than those in the ACG. The average number of inpatient days was lower and breast cancer screening rates were higher for IG members than for ACG members for Vendor A's Local sample. Since we determined that these impacts were not driven by participation in the treatment arms, we can conclude that placebo effects do exist for these outcomes. Assignment to the IG was associated with lower total Medicare spending, Medicare inpatient spending, and fewer inpatient days for Vendor A's Local sample. We did not find any placebo effects for Vendor A's National sample or for either of Vendor B's samples.

Table 17. Vendor A Local, Placebo Effects[a]

	Coefficient on Pooled Arms 1, 2	P-value	Coefficient on IG	P-value	Potential Placebo Effect
Medicare Payment					
Total Medicare payment	103	0.860	-756***	0.008	Yes
Inpatient payment	-195	0.562	-478***	0.005	Yes
Outpatient payment	197	0.178	21	0.786	No
Carrier payment	10	0.922	-75	0.193	No
ED payment	12	0.290	-4	0.360	No
Office visit payments	-1	0.979	-31	0.473	No
Medicare Use					
Any inpatient visits	0.8%	0.667	-0.9%	0.264	No
Inpatient days	-0.1	0.659	-0.3***	0.001	Yes
Any outpatient visits	2.3%	0.136	0.4%	0.522	No
Office days	0.0	0.998	-0.2	0.297	No
Any ED visits	3.4%	0.124	-0.7%	0.502	No
Preventive Screening					
Colorectal cancer	0.7%	0.756	0.9%	0.374	No
Breast cancer	-1.8%	0.455	2.1%*	0.063	Yes
Cardiovascular disease	-1.6%	0.469	0.3%	0.703	No

[a]Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

4.3.6 Regression to the Mean

In order to gain additional insight into the results presented in this chapter, we assess the effect of the regression-to-the-mean process. Regression to the mean can cause spurious results in studies that use repeated measures but whose treatment and comparison groups differ in their outcomes' distributions at baseline. For example, if pre-period expenditures for the control group were very low while those of the treatment group were very high, it is likely that we would observe a large negative impact of the demonstration because post-period expenditures would likely increase for the control group and decrease for the treatment group as a result of the regression-to-the-mean process.

In particular, we consider total Medicare expenditures per beneficiary in the ACG during the pre- and post-implementation periods. We focus on the control group to ensure that the post-period means are not influenced by participation in the demonstration. (As described earlier, individuals in the ACG were randomly selected and were not contacted by vendors). Table 18 presents information on pre- and post-period costs for the ACG. Columns 5 and 6 indicate that changes in pre- versus post-period costs were negatively related to cost category during the pre-demonstration period. For example, the lowest-cost pre-period beneficiaries (having average expenditures of $851) experienced an average increase in total expenditures of $3,140

in the post period whereas the highest cost beneficiaries (averaging $45,832 in the pre-period) experienced a decrease of $24,158 in the post period.

Table 18: Regression to the Mean among ACG Beneficiaries

Pre-period Cost Category ($)	N	Mean Cost per Beneficiary, Pre-Period ($)	Mean Cost per Beneficiary, Post-Period ($)	Change in Mean Pre- vs. Post-Period Cost ($)	Change in Mean Pre- vs. Post-Period Cost (%)
(1)	(2)	(3)	(4)	(5)	(6)
<2,500	27,930	851	3,991	3,140	369.2
2,500-5,000	7,223	3,537	7,206	3,669	103.7
5,000-7,500	2,925	6,115	9,204	3,089	50.5
7,500-10,000	1,581	8,673	10,893	2,220	25.6
10,000-12,500	1,149	11,192	11,421	229	2.0
12,500-15,000	786	13,705	11,123	-2,582	-18.8
15,000-17,500	660	16,155	13,401	-2,754	-17.0
17,500-20,000	522	18,716	14,150	-4,567	-24.4
20,000-22,500	412	21,117	14,744	-6,373	-30.2
22,500-25,000	342	23,705	14,357	-9,348	-39.4
>=25,000	1,905	45,832	21,674	-24,158	-52.7
Overall mean		5,042	6,560	1,518	30.1

Since the SRRD randomly selected beneficiaries for the treatment and comparison groups, such a "mismatch" should not pose a threat to the evaluation. However, as discussed above, there were some differences in baseline means for Vendor B. Thus, for Vendor B we reran the regressions for those individuals with lower than average baseline total Medicare expenditures and separately for those with higher than average baseline total Medicare payments. The results did not change, implying that regression to the mean is not driving our findings.

4.4 Discussion and Conclusions

The existing literature has found positive impacts of HRAs with tailored follow-up risk reduction programs when implemented among the working age population. The application of these programs to Medicare beneficiaries provides a unique opportunity to assess whether similar programs would work among a population with high risk and multiple chronic conditions. The design of this demonstration and the availability of the Medicare claims data allow for the investigation of impacts that are generally difficult or impossible to identify in wellness and prevention programs that rely on voluntary participation.

For each vendor, we examined the impacts of random assignment into treatment arms, in which participants received the tailored risk reduction interventions and assignment into the "HRA only" (a control arm where participants received only a general brochure on good health practices). We examined health expenditures, health services use, and certain preventive screening measures constructed using Medicare administrative claims data during the demonstration and a baseline period before the start of the demonstration.

We did not find any impacts of SRRD on outcomes examined for the overall demonstration for either vendor. We did find variation in the impact of the SRRD by demonstration year. For Vendor A, the demonstration did decrease total Medicare expenditures and hospital inpatient expenditures in Year 1. For Vendor B, we found lower total Medicare expenditures and Medicare use, and higher preventive screening rates among the vendor's Local sample in Year 2. However, for both Vendor A and B these impacts did not persist into later demonstration years.

A limitation of this study is that data were available only for the three demonstration years. It is likely that any effect of SRRD on behavior will take longer to manifest in the health claims data through changes in spending and utilization than the relatively short period of this study. It is also possible that the program duration may need to be longer for behavior changes to manifest in changes in health expenditures and utilization. Thus, it is important to consider the pathway by which the SRRD would improve health.

5. IMPACT OF SRRD ON HRA-BASED HEALTH OUTCOMES

5.1 Introduction

In this chapter, we examine the impact of SRRD on health outcomes constructed from responses to the HRAs administered each year by the vendors. Based on responses to HRA questions, each vendor applied proprietary risk stratification algorithms that categorized beneficiaries into high, medium, and low risk for risk areas related to such topics as physical activity, nutrition, preventive care, and depression. We present the impact of the SRRD on two types of HRA-based outcomes: (1) participants' risk scores constructed by the vendors using participants' responses and (2) self-reported health status measures. To examine SRRD impacts, we compared these HRA-based outcomes between the treatment group of beneficiaries receiving the risk reduction intervention and the control group beneficiaries receiving only the HRA and a generic health brochure.

Summary of Results

- For Vendor A, participation in the risk reduction programs resulted in lower overall risk levels and lower probability of being high risk in the following areas: stress and general well-being.
- For Vendor A, participation also resulted in higher self-rated health status.
- Participation in Vendor B's risk reduction programs resulted in lower overall risk levels and lower probabilities of being high risk in the following areas: back care, nutrition, physical activity, and general well-being. Those in participated in risk reduction programs for all three years also showed improvements in the stress risk area.
- Vendor B's risk reduction program participants also resulted in higher self-rated health status.
- For Vendor B, we were able to assess whether longer participation in the risk reduction programs resulted in larger impacts. Beneficiaries who participated for all three years of the demonstration did indeed have the largest decreases in the probability of being categorized as high risk in the same areas listed above, as well as greater improvements in self-rated health status.
- SRRD had no impact on the risk areas of motor safety, weight, polypharmacy and falls.

One challenge in assessing the impact of SRRD on HRA outcomes is that beneficiaries selected whether or not to re-enroll each year. As a result, yearly HRA information was only available for individuals choosing participate (i.e., re-enroll) each year. Re-enrollees tended to be white, less likely to be dual eligible, and more likely to obtain certain preventive screenings. Thus, the group of beneficiaries with multiple HRAs was not the same as the group who chose not to participate after Year 1.

To address this challenge, we used a two-step multivariate regression method designed to address sample selection issues. Our focus is on assessing whether receiving the SRRD intervention was effective in reducing risk and improving self-reported health status.

5.2 Data and Methodology

5.2.1 SRRD Design

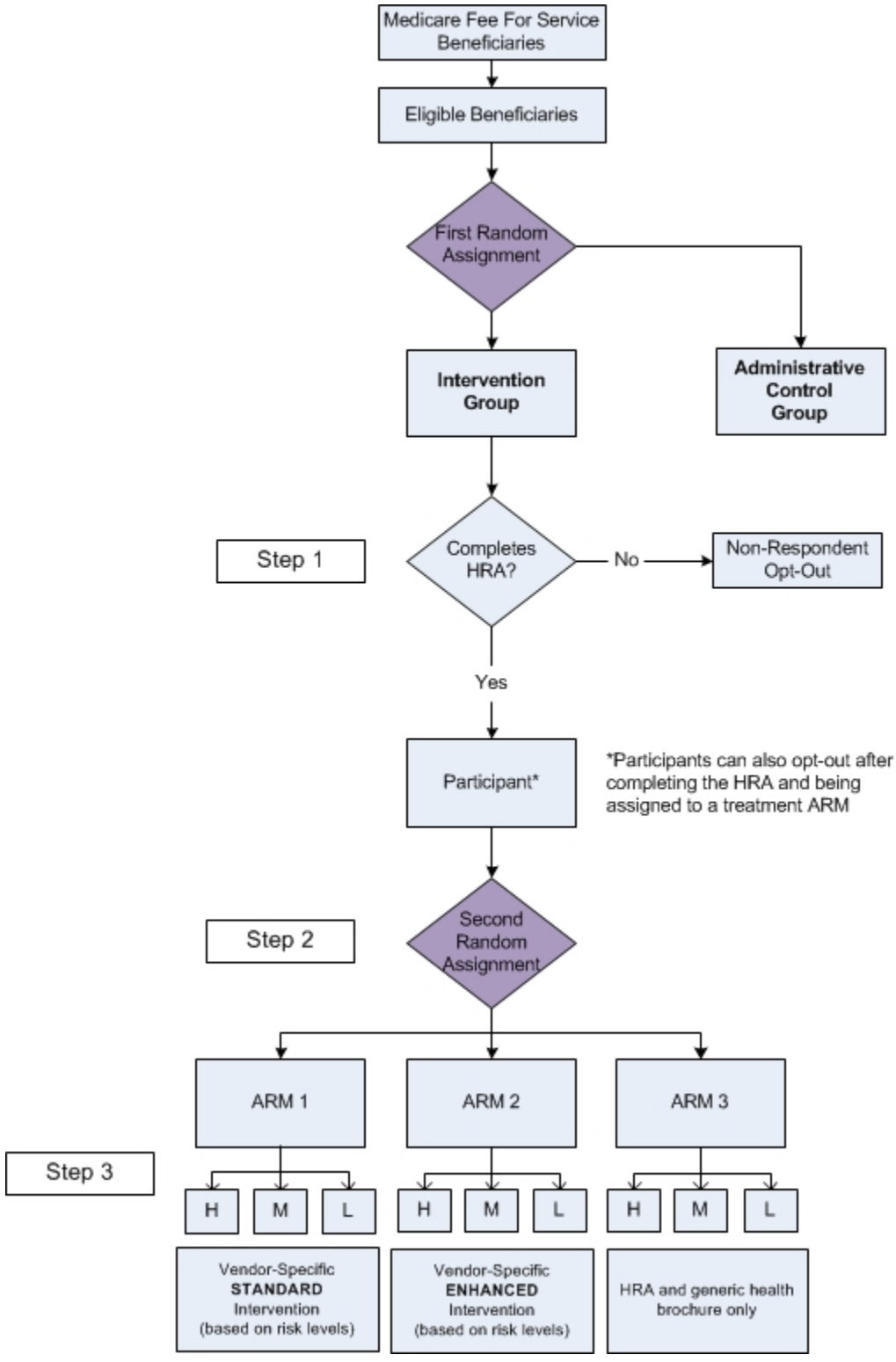
The SRRD demonstration had a RCT design with two rounds of random assignment. For each vendor, 40,000 eligible Medicare beneficiaries were randomly assigned to an Intervention Group (IG) or an Administrative Control Group (ACG) (20,000 in each group).[52] Each vendor then mailed a health risk assessment (HRA) packet to its assigned IG beneficiaries to recruit them for the study (i.e., invite them to participate). Figure 1 provides an overview of the SRRD design. Once a vendor received a completed HRA from a beneficiary, the beneficiary was considered successfully recruited into the study and became an SRRD participant (Step 1 from Figure 1). Participants were then randomized a second time into one of three intervention arms for each vendor (Step 2 from Figure 1):

- Arm 1 – Standard Treatment (HRA + standard tailored follow-up)
- Arm 2 – Enhanced Treatment (HRA + enhanced tailored follow-up)
- Arm 3 – HRA Only (HRA + generic health advice).

Arms 1 and 2 participants received varying levels of tailored intervention services aimed at reducing risk and improving health, depending on the treatment Arm to which they were randomly assigned, and each beneficiary's HRA responses. Arm 3 participants received only a generic health mailing. The SRRD demonstration started May 1, 2009 and ended April 31, 2012.

[52] Vendors received a "refresh" sample consisting of an additional 11,800 individuals for each vendor at the end of the first year of the demonstration.

Figure 1. SRRD Study Design



Note: H = High Risk | M = Moderate Risk | L = Low Risk

5.2.2 HRA Description

Each vendor independently used previously developed HRA questionnaire tools to design customized HRAs for the SRRD. The HRAs gathered beneficiary information on the following topics:

- Self-rated health status
- Chronic conditions (e.g., arthritis, cancer, high blood pressure, kidney disease)
- Obesity
- Blood pressure
- Cholesterol
- Lifestyle (e.g., physical activity level, diet, tobacco use, and alcohol use)
- Stress and emotional well being
- Prescription medication use
- Preventive care (e.g., primary health care provider, flu shot, last physical examination)
- Perceived importance of and confidence in making healthy lifestyle changes.

Based on responses to the HRA questions, the vendors applied proprietary risk stratification algorithms to categorize beneficiaries into high, medium, or low risk for risk factors. Based on these categorizations and an overall risk score, vendors tailored follow-up health risk reduction programs depending on the beneficiary's treatment arm and risk categorization. The follow-up programs consisted of a tailored feedback report based on the HRA; health coaching via mail, Internet, or phone; and links to community resources. Those assigned to Arm 2 received more intense health coaching through more frequent contact.

Note that the two vendors' HRA questionnaires and risk stratification algorithms differed. Thus, we caution against comparing risk topics across the two vendors (e.g. although both vendors address stress, the questions and responses used to generate a risk categorization for stress differ across the two vendors).

5.2.3 HRA Data Collection

Figure 2 shows the HRA collection schedule for the demonstration. Vendors collected HRAs at the beginning of each demonstration year, which began on May 1 and ended on April 30 of the following year. Since Vendor A exited the demonstration at the end of Year 2, only two rounds of HRAs were available for Vendor A's participants (Year 1 and Year 2). Vendor B remained for all three years of the demonstration. Exit HRAs were available for Vendor B in addition to HRAs for Years 1, 2, and 3.

Figure 2. Timing of HRA Collection

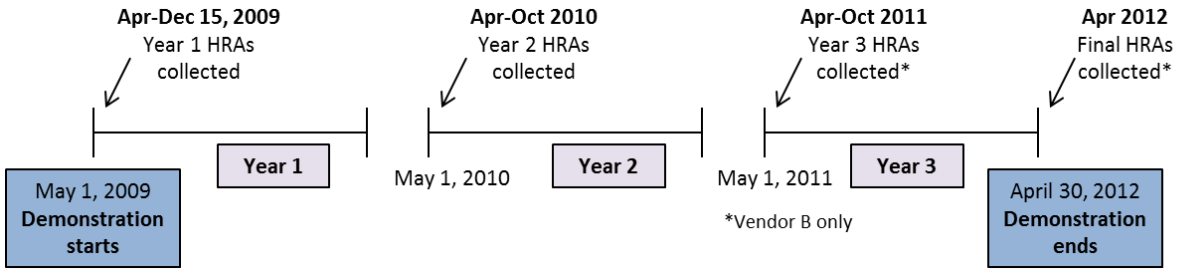


Table 1 shows the number of HRAs collected by Vendor A by treatment arm for each year. Overall, Vendor A collected 12,145 HRAs in the two years Vendor A was in the demonstration. These HRAs were collected from 7,988 participants across all arms. The last column of Table 1 reports the percentage of Vendor A's participants completing HRAs in both years. Overall, 52 percent of participants completed two HRAs, with the highest rates of multiple HRA completion for those assigned to the HRA-only arm, or Arm 3 (62 percent).

Table 2 shows that Vendor B collected a total of 20,046 HRAs over the three years of the demonstration. These HRAs were collected from 7,773 participants. Table 2 also shows the percentage of participants completing Year 1 and Year 2 HRAs, the percentage of completing HRAs in at least any two years and the percentage of participants completing HRAs in all three demonstration years. Overall, 50 percent of participants completed both Year 1 and Year 2 HRAs, 66 percent completed HRAs in at least two years, and 37 percent completed HRAs in all three years. Participants assigned to Arm 3 had the highest completion rates. Later in this chapter, we discuss the implications of the differences in HRA completion rates on our analyses.

Table 1. Vendor A HRA Completions by Arm[a]

Treatment Arm	Year 1 HRA	Year 2 HRA	Total HRAs	Percent Participants-Year 1 & 2 HRAs[b]
Arm 1. Standard	2,230	1,719	3,949	48%
Arm 2. Enhanced	2,232	1,683	3,915	47%
Arm 3. HRA-only	2,220	2,061	4,281	62%
All Arms	6,682	5,463	12,145	52%

[a] Excludes HRAs of a "refresh" sample provided to vendors at the end of Year 1/beginning of Year 2, as well as participants who died or moved to Medicare Part C during the demonstration.

[b] The percentage of participants with multiple HRAs is calculated by dividing the number of participants submitting HRAs in both Year 1 and Year 2 by the number of participants submitting an HRA in Year 1 or Year 2.

Table 2. Vendor B HRA Completions by Arm[a]

Treatment Arm	Year 1 HRA	Year 2 HRA	Year 3 HRA	Exit HRA	Total	Percent Participants, Year 1 & 2[b]	Percent Participants, ≥ 2 Years[c]	Percent Participants, 3 Years[d]
Arm 1. Standard	2,109	1,654	1,176	1,259	6,198	47%	61%	31%
Arm 2. Enhanced	2,134	1,610	1,152	1,259	6,155	45%	59%	29%
Arm 3. HRA-only	2,151	1,979	1,809	1,754	7,693	59%	77%	51%
All Arms	6,394	5,243	4,137	4,272	20,046	50%	66%	37%

[a] Excludes HRAs of a "refresh" sample provided to vendors at the end of Year 1/beginning of Year 2, as well as participants who became ineligible during the demonstration.

[b] The percentage of participants with Year 1 & 2 HRAs is calculated by dividing the number of participants submitting HRAs in both Year 1 and Year 2 by the number of participants submitting an HRA in Year 1 or Year 2.

[c] The percentage of participants with HRAs for at least two years is calculated by dividing the number of participants submitting HRAs in at least any two years of Year 1, Year 2, or Year 3 by the number of participants submitting an HRA in Year 1, Year 2, or Year 3

[d] The percentage of participants with HRAs with three years is calculated by dividing the number of participants submitting HRAs in Year 1, Year 2, and Year 3 by the number of participants submitting an HRA in at least Year 1, Year 2, or Year 3.

5.2.4 HRA Risk Variables

Outcome values for Vendor A are summarized in Table 3 and outcome values for Vendor B in Table 4. The first set of outcomes listed in the tables comprise topic-specific risk scores that vendors constructed based on responses to specific HRA questions. The remaining two outcomes are the overall risk score and a measure of self-reported health status. Because the vendors used proprietary algorithms to generate the topic-specific and overall risk scores, the definition and scales for each outcome varied by vendor and are not directly comparable between vendors.

Table 3. Vendor A Selected Risk Variables and Yearly Means

Variable	Year 1	Year 2
Excessive alcohol use (% high risk)	2.4%	2.0%
Falls (% high risk)	19.8%	20.1%
Poor nutrition (% high risk)	18.6%	15.6%
Physical inactivity (% high risk)	21.0%	19.7%
Inappropriate use of clinical preventive services (% high risk)	1.6%	1.3%
Depression (% high risk)	28.8%	27.3%
Polypharmacy/medication issues (% high risk)	4.0%	4.1%
High stress (% high risk)	10.0%	9.2%
Smoking/tobacco use (% high risk)	8.7%	7.8%
Overweight/obese (% high risk)	29.3%	28.7%
Lack of general well-being (% high risk)	16.4%	15.7%
Overall risk points mean (0 = low, 17 = high)	6.4	6.2
Self-rated health status mean (1=poor, 5=excellent)	3.3	3.3

Table 4. Vendor B Selected Risk Variables and Yearly Means

Variable	Year 1	Year 2	Year 3	Exit HRA
Excessive alcohol use (% high risk)	3.9%	3.6%	3.1%	2.9%
Poor back care (% high risk)	40.4%	39.6%	39.1%	37.7%
Lack of motor safety (% high risk)	1.5%	1.2%	1.1%	0.9%
Poor nutrition (% high risk)	21.0%	19.4%	19.8%	19.1%
Physical inactivity (% high risk)	22.0%	20.9%	20.3%	20.5%
Inappropriate use of clinical preventive services (% high risk)	3.2%	1.2%	1.1%	0.8%
High stress (% high risk)	5.9%	5.4%	5.9%	5.6%
Smoking/tobacco use (% high risk)	2.1%	1.6%	1.3%	1.4%
Overweight/obese (% high risk)	28.1%	27.3%	25.2%	25.4%
Lack of general well-being (% high risk)	10.0%	10.3%	10.1%	9.1%
Overall risk (0=low to 100=high)	33.1	32.3	31.8	31.5
Self-rated health status (1=very poor to 6=excellent)	4.5	4.5	4.5	4.5

Tables 3 and 4 also provide the mean value or percentage of individuals categorized as high risk each year, depending on the measure. Vendors assessed participants as high, medium, and low risk. We combined medium and low into one category and reported the percentage of participants categorized as high risk. As noted above, each vendor defined risk factors and risk categories differently. For example, Vendor A's overall risk variable ranged from 0 (low risk) to 17 (high risk) while Vendor B used a 0 (low risk) to 100 (high risk) scale.

Table 3 shows the percentage of participants who were high risk in each year for each topic-specific variable for Vendor A. In Year 1, 2.4 percent of Vendor A's participants were categorized as high risk for excessive alcohol use; in Year 2, 2.0 percent were categorized as high risk for excessive alcohol use. The last two rows show the mean overall risk and the mean self-rated health status (ranging from 1 = poor to 5 = excellent). The means between Year 1 and Year 2 appear stable, with Year 2 means generally lower than Year 1 means (indicating improvements in risk). However, fewer individuals participated in Year 2 and the participants who did choose to re-enroll for a second year differed from those who did not choose to re-enroll.

Table 4 shows the percentage of participants who were high risk in each year and for the Exit HRA for Vendor B. The last two rows show the mean overall risk (ranging from 0 = low to 100 = high) and the mean self-rated health status (ranging from 1 = very poor to 6 = excellent). The percentage of high risk individuals also generally decreased each year, though self-reported health status was unchanged.

5.3 Methodology

In this section, we describe our methodology for examining the impact of the SRRD on these HRA outcomes. To examine the impact of the demonstration on participant risk levels and self-reported health status we assessed whether Arms 1 and 2 participants (those receiving treatment) had improved outcomes compared with Arm 3 participants who did not receive the follow-up risk reduction intervention (see Figure 1).

We needed to observe at least two HRAs for each participant—a baseline HRA and at least one follow-up HRA—to detect changes that may be attributed to the SRRD. One approach is to limit the sample to beneficiaries who completed at least two HRAs. However, as shown in Table 1, limiting the sample in this way would result in dropping a large percentage of participants. This in itself would not pose a significant hurdle to the analyses with respect to sample size, since the remaining sample would be sufficient large. The more important difficulty resulting from keeping only re-enrollees is that attrition from the demonstration each year was not random—participants chose to drop out of the demonstration or re-enroll for another year based on factors that are not fully observable with the data available.

As shown in Tables 1 and 2, attrition from the demonstration occurred each year and was not randomly distributed across the treatment arms—fewer individuals dropped out of Arm 3 (the HRA-only arm) than out of Arms 1 and 2. For Vendor A, for example, 48 percent and 47 percent of Arm 1 and Arm 2 participants, respectively, completed an HRA in Year 2 versus 62 percent of Arm 3 participants. The completion rate pattern for Vendor B was similar.

Tables 5 and 6 show that high-/medium-risk individuals in Year 1 were less likely to re-enroll and continue to participate than low-risk individuals for both vendors across each of the arms. For Vendor A, 54 percent of high- or medium-risk individuals re-enrolled into Year 2 of the demonstration versus 67 percent of low-risk individuals. For Vendor B, the re-enrollment rates also varied by Year 1 risk levels. Individuals assigned to Arm 3 had higher re-enrollment rates among both low-risk and high-/medium-risk participants compared with Arm 1 and Arm 2. For Vendor A, 75 percent of low-risk participants in Arm 3 re-enrolled compared with 62 percent of low-risk participants in Arm 2. For high-risk participants, 72 percent of those in Arm 3 re-enrolled while only 43 percent of high-risk participants in Arm 2 re-enrolled. The patterns were similar for Vendor B.

Table 5. HRA Completion Rates by Year 1 Overall Risk Levels

Treatment Arm	Year 1 Risk Categorization[a]	Vendor A	
		Year 1	Year 2[b]
Arm 1 (Standard)	High/Medium Risk	100%	46%
	Low Risk	100%	63%
Arm 2 (Enhanced)	High/Medium Risk	100%	43%
	Low Risk	100%	62%
Arm 3 (HRA-Only)	High/Medium Risk	100%	72%
	Low Risk	100%	75%
All Arms	High/Medium Risk	100%	54%
	Low Risk	100%	67%

[a]We combined high and medium risk instead of medium- and low-risk due to limited availability of Vendor B's algorithm for determining overall risk cutoffs.

[b]Completion rates in Year 2 were calculated by dividing the number of high- or medium-risk participants in Year 1 who completed an HRA in Year 2 by the number of high- or medium-risk participants in Year 1.

Table 6. HRA Completion Rates by Year 1 Overall Risk Levels

Treatment Arm	Year 1 Risk Categorization[a]	Vendor B		
		Year 1	Year 2[b]	Year 3[c]
Arm 1 (Standard)	High/Medium Risk	100%	55%	43%
	Low Risk	100%	62%	55%
Arm 2 (Enhanced)	High/Medium Risk	100%	51%	42%
	Low Risk	100%	62%	52%
Arm 3 (HRA-Only)	High/Medium Risk	100%	70%	70%
	Low Risk	100%	74%	75%
All Arms	High/Medium Risk	100%	58%	51%
	Low Risk	100%	66%	61%

[a]We combine high and medium risk instead of medium and low risk due to limited availability of Vendor B's algorithm for determining risk cutoffs.
[b]Completion rates in Year 2 were calculated by dividing the number of high- or medium-risk participants in Year 1 who completed an HRA in Year 2 by the number of high- or medium-risk participants in Year 1.
[c]Completion rates in Year 3 were calculated by dividing the number of high- or medium-risk participants in Year 1 who completed an HRA in Year 3 by the number of high- or medium-risk participants in Year 1.

The non-random attrition indicates that our methodology for assessing the impact of SRRD on HRA outcomes must account for re-enrollment selection by participants. If we simply included participants with completed HRAs, as noted, our results might be confounded by the fact that risky, less healthy participants chose to remain within the program for Arm 3 and not for Arms 1 and 2, leading to potential incorrect attribution of improvements in risk levels to the demonstration.

For this reason, we used the Heckman Selection Model to control for potential selection bias.[53] This model incorporates a two-stage estimation process. The first stage estimates the probability of participating in the demonstration for at least two years for each individual. The probability of re-enrollment is expressed as a continuous unobserved variable, P_i^*, which is only observed if $P_i^* > 0$ (Equations 1 and 2):

Equation 1: $$P_i^* = \beta_0 + \beta_1 Z_i + \epsilon_i$$

Equation 2: $$P_i = \begin{cases} 1 \text{ if } P_i^* > 0 \\ 0 \text{ if } P_i^* \leq 0 \end{cases}$$

The second stage consists of regressing outcome measures (R_i) on explanatory variables conditional on participation and a correction for self-selection by incorporating the probabilities estimated from the first stage (Equations 3 and 4):

[53] Heckman, James. "Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47, 153-161, 1979.

Equation 3. $R_i = \delta_0 + \delta_1 X_i + \mu_i$ if $P_i = 1$

Equation 4. $R_i =$ Unobserved if $P_i = 0$

The Heckman Selection Model relies on the availability of exclusion restrictions that directly explain re-enrollment but only affect the outcome measure indirectly through impacting re-enrollment. We use the following exclusion restrictions for Vendors A and B:

- Vendor A: *timing of HRA collection*—Vendor A staggered the recruitment of beneficiaries in six waves of recruitment mailings spread over several months. Beneficiaries contacted earlier versus later may be more or less likely to participate based on the time of year they were contacted. This would not be expected to impact HRA outcomes. Vendor B did not stagger its HRA administration in the first year. In later years Vendor B did transition to a wave approach, but wave information was not available in the data.
- Vendors A and B: *mode of HRA completion*—Beneficiaries had the option to complete the HRA via a paper or online method. The majority of participants completed the HRA via paper. However, the ease of one method versus the other may have influenced the beneficiaries' decision to participate in the following year. We acknowledge that those selecting to complete the HRA online may differ in some systematic way from those using the paper method, which may also be associated with behaviors or characteristics that affect the outcomes interest. Although we cannot directly test for this possibility, we did examine whether the mode of HRA completion was statistically significant in explaining outcomes among participants when controlling for treatment arm assignment. We did not find this to be the case, suggesting that the mode of HRA completion can be a valid exclusion restriction.

We estimated the two-stage model using the methodology described above separately for each vendor and for each of the outcomes listed in Tables 3 and 4. The first stage was estimated using a probit model including as explanatory variables the exclusion restrictions listed above as well as the following variables:

- Participation in Arms 1 or 2 (indicator variable for receiving the intervention)[54]
- Baseline outcome values (for Vendor A, the baseline is always Year 1; for Vendor B, the baseline is the earliest HRA completed by the individual)
- Demographic characteristics
  - Age
  - Female (indicator variable)
  - Nonwhite (indicator variable)
- Medicare-Medicaid dual eligibility (indicator variable)

[54] In the analyses presented in this chapter, we pool individuals in Arms 1 and 2 in order to increase sample size.

- Two or more chronic conditions (indicator variable)
- Total Medicare payments from the baseline year.

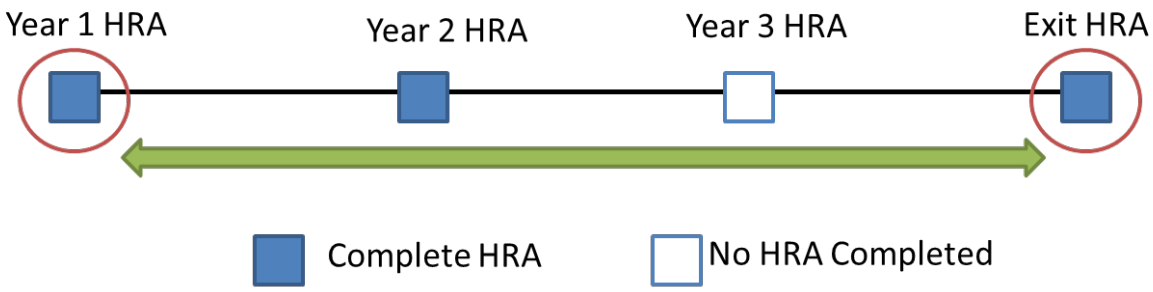
In the second stage, we use the same explanatory variables with the exception of the exclusion restrictions. The coefficient on the indicator for Arm 1 or 2 in this second stage provides an estimate of the impact of participation in the demonstration on each of the outcome measures.

For each outcome, we examined the impact of the demonstration on three types of participants:

1. Individuals who re-enrolled in Year 2 (Vendors A and B, separately)
2. Individuals who re-enrolled at least once in any year (Vendor B only)[55]
3. Individuals who participated for the full length of the demonstration (Vendor B only).

For participant type #2, we measured the impact of SRRD on individuals as the difference between the last HRA completed and the first HRA completed. For example, if an individual completed Year 1, Year 2, and Exit HRAs, we examine the change in outcomes between the Exit HRA and Year 1 HRA. This scenario is illustrated in Figure 3, which shows the case of a Vendor B participant who completed Year 1, Year 2, and the Exit HRA. The circles around the Year 1 and exit HRAs and the arrow connecting the Year 1 and Exit HRAs indicate that these two data points are included in the analysis.

Figure 3. Impact Analysis for Individuals Re-enrolling in Any Year



We also ran all the models using an alternative non-response weighting method, in which the probably of re-enrolling is estimated using a logistic regression in the first stage. The predicted values of re-enrollment are then used to weight the impact regression on outcomes in the second. This method yielded very similar results.

[55] Note that for Vendor A this analysis would be exactly the same as #1, since Vendor A only participated for Years 1 and 2 of the demonstration. For Vendor B, individuals who re-enrolled at least once in any year are a subset of the individuals who re-enrolled in Year 2 (i.e., type #1).

5.4 Findings

Table 7 reports estimated impacts of SRRD for each of the HRA-based outcomes for Vendor A. Impacts are reported as marginal effects of Arm 1 or Arm 2 participation (pooled) compared with Arm 3 participation. Each outcome variable represents a separate regression (the estimated coefficients on other explanatory variables for each regression are suppressed).

For Vendor A, Table 7 shows that pooled Arms 1 and 2 participants had a lower overall risk score (statistically significant at the 1 percent level) and a higher self-rated health status compared with Arm 3 participants (statistically significant at the 10 percent level). The following results were statistically significant at the 1 percent level:

- Lower probability of being high risk for stress, and
- Lower probability of being high risk for overall well-being.

Table 7. Vendor A Effect of the Demonstration on Year 2 HRA-based Outcomes

Outcome Variable[a]	Marginal Effect Pooled Arms 1&2[b]
Overall risk (0=low to 17=high)	-0.571***
Self-rated health status (1=poor to 5=excellent)	0.059*
Excessive alcohol use (high risk)	0.001
Falls (high risk)	-0.015
Poor nutrition (high risk)	-0.012
Physical inactivity (high risk)	-0.012
Inappropriate use of clinical preventive services (high risk)	0.072
Depression (high risk)	-0.033
Polypharmacy/medication issues (high risk)	0.014
High stress (high risk)	-0.015***
Smoking/tobacco use (high risk)	-0.001
Overweight/obese (high risk)	0.002
Lack of general well-being (high risk)	-0.030***

[a] Regression results were obtained by using a Heckman Selection Model, in the first stage to estimate the probability of participating in the Year 2 of the demonstration and in the second stage to estimate the overall risk level, self-rated health status, and probability of being high risk in a series of risk areas.

[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

Table 8 presents the results for Vendor B. In the second column, we present the marginal effects of pooled Arms 1 and 2 participants on each of the outcome variables in Year 2; in the third column, we present the results for individuals submitting HRAs in any 2 years; in the fourth column we report results for individuals participating in all three years. For the three

analyses, pooled Arms 1 and 2 participants showed lower overall risk scores, higher self-rated health status, lower probabilities of poor nutrition risk, physical inactivity risk, poor back care and lack of general well-being risk (all statistically significant at the 1 percent level). Pooled Arms 1 and 2 individuals who participated for at least two years had lower risk of excessive alcohol use (statistically significant at the 1 percent level) and smoking/tobacco use (statistically significant at the 10 percent level), but these results were not found for those participating in all three years. Arm 1 and 2 individuals participating for all three years showed improved risk for high stress (statistically significant at the 5 percent level).

For those who participated in Arms 1 or 2 in all three years of the demonstration, the magnitude of the decrease in overall risk was larger than for those who did not participate over the full demonstration period.

Table 8. Vendor B Impact of SRRD on HRA-based Outcomes

Outcome Variable[a]	Marginal Effect, Pooled Arms 1&2 (Year 2)[b]	Marginal Effect, Pooled Arms 1&2 (Any Year)[b]	Marginal Effect, Pooled Arms 1&2 (Full)[b]
Overall Risk (0=low to 100=high)	-0.948*	-1.030***	-2.206**
Self-rated health status (1 = very poor to 6 = excellent)	0.166***	0.200***	0.342***
Excessive alcohol use (high risk)	-0.014***	-0.008**	-0.007
Poor back care (high risk)	-0.039	-0.058***	-0.082***
Lack of motor safety (high risk)	-0.003	-0.003	-0.003
Poor nutrition (high risk)	-0.028***	-0.042***	-0.058***
Physical inactivity (high risk)	-0.030***	-0.037***	-0.045***
High stress (high risk)	0.007	-0.007	-0.016**
Smoking/tobacco use (high risk)	-0.007**	-0.005*	-0.003
Overweight/obese (high risk)	0.006	-0.009	-0.010
Lack of general well-being (high risk)	-0.016***	-0.032***	-0.036***

[a] Regression results are obtained by using a Heckman Selection Model, in the first stage to estimate the probability of participating in Year 2 of the demonstration and in the second stage to estimate the overall risk level, self-rated health status, and probability of being high risk in a series of risk areas.

[b] Statistical significance is shown at the 1% (***) , 5% (**) and 10% (*) levels.

We also estimated the impacts of each of the arms separately (Arm 1 versus Arm 3 and Arm 2 versus Arm 3) and did not find the results to change substantially. Consistent with this finding, we did not find any additional impact of Arm 2 over Arm 1, though sample sizes become an issue when looking at this smaller subgroup.

5.5 Conclusion and Discussion

We used information from the HRAs to examine the impact of the SRRD on risky health behaviors and self-rated health status, dimensions that the Medicare claims data do not capture. SRRD-induced improvements in behavior might take years to translate into changes in costs and use that could be captured in the health claims data. The HRAs potentially provide insight into the pathways through which the SRRD operated. It would be difficult to identify these patterns in the health claims data even with a detailed and resource-intensive investigation into diagnosis and procedure codes. Comparing individuals' HRA data pre- and post-intervention, in contrast, allows us to investigate whether the SRRD caused changes in health risk behaviors, which are likely linked to improved health outcomes in the longer run.

The SRRD was successful in reducing health risk in certain areas. Specifically, beneficiaries participating in Vendor A's risk reduction programs were less likely to be high risk for stress and lack of general well-being. We also detected a decrease in overall risk for Vendor A's participants. Beneficiaries participating in Vendor B's risk reduction programs were less likely to be high risk for poor back care, poor nutrition, physical inactivity, and general well-being, as well as overall risk. Both vendors' participants showed improvements in self-rated health status. The greatest improvements in risk were seen for those who participated for the full three years. For those that participated for the full three years, results showed reductions in being high risk for stress as well as in the areas of poor back care, poor nutrition, physical inactivity and general well-being.

6. FINAL BUDGET NEUTRALITY RESULTS

6.1 Introduction

In this section we present final budget neutrality (FBN) findings for the SRRD. The purpose of the FBN calculations is to provide CMS with the information necessary to conduct a reconciliation of total payments to SRRD vendors. CMS committed to paying vendors for each recruited beneficiary (i.e., participant) in the demonstration; however, CMS deferred 10 percent of payments due to vendors until after demonstration completion.[56] The FBN analysis compares each vendor's Medicare expenditures for the intervention group (IG) to Medicare expenditures for the administrative control group (ACG) to determine whether each vendor's SRRD program was budget neutral. A vendor that was not budget neutral would receive either a partial payment of the deferred fees or none, depending on the budget neutrality results. Budget neutrality and the return of fees were determined as follows:

> *Budget neutral, vendor receives deferred fees:* If the Medicare expenditures for the IG were lower than those of the ACG by at least the amount of the demonstration payments made to the vendor plus deferred fees, CMS would return all deferred fees to the vendor.
>
> *Not budget neutral, vendor receives partial fees*: If the Medicare expenditures for the IG were lower than those for the ACG by enough to cover the demonstration payments made to the vendor, but not all the deferred fees, CMS would return the difference between the savings and the deferred fees.
>
> *Not budget neutral, vendor receives no deferred fees*: If the Medicare expenditures for the IG were not lower than those for the ACG or not low enough to cover the demonstration payments made to the vendor, CMS would not return any deferred fees.

We present the findings for the following three FBN methodologies:

> Method 1. An FBN methodology that uses a *simple differences method* to compare the Medicare IG and ACG spending for each vendor in the 36 months of the demonstration period.
>
> Method 2. An FBN methodology that considers any divergences in baseline expenditures by using a *difference-in-differences (DID)* comparison of the Medicare IG and ACG spending for each vendor in the 36 months of the demonstration *and* the 16 months of the baseline period before the start of the demonstration.

[56] Based on the results of an interim budget neutrality (IBN) analysis conducted after the first 15 months of the demonstration, CMS maintained a deferral rate of 10 percent for Year 3 of the demonstration, instead of increasing the deferral rate to 20 percent (or accepting a vendor's resignation from the demonstration for Year 3).

Method 3. An FBN methodology that controls for differences in baseline expenditures and some beneficiary characteristics, by using a *regression-based adjustment* to compare the Medicare IG and ACG spending for each vendor.

Table 1 and Table 2 summarize the results for Vendor A and Vendor B, respectively. For Vendor A, all three methods yield the same result – Vendor A is budget neutral. For Vendor B, method 2 indicates that Vendor B is budget neutral while Method 1 and Method 3 indicate that Vendor B is not budget neutral.

Table 1. Vendor A Summary of FBN Results

	Method 1 Simple Differences	Method 2 Difference in Differences	Method 3 Regression Adjustment
FBN Results	Budget neutral	Budget neutral	Budget neutral
Reconciliation	CMS pays total deferred fees	CMS pays total deferred fees	CMS pays total deferred fees

Table 2. Vendor B Summary of FBN Results

	Method 1 Simple Differences	Method 2 Difference in Differences	Method 3 Regression Adjustment
FBN Results	Not budget neutral	Budget neutral	Not budget neutral
Reconciliation	CMS keeps total deferred fees	CMS pays total deferred fees	CMS keeps total deferred fees

In the next sections, we provide step-by-step descriptions of all three FBN methodologies and present the findings.

6.2 Final Budget Neutrality Methodologies

In calculating total Medicare expenditures for the IG and ACG, we considered adjusting expenditures for the number of months of eligibility of IG and ACG members, but did not do so because:

(1) The monthly eligibility data were not available and alternative approaches would have required approximating the timing of eligibility changes.

(2) The intent-to-treat study design yields unbiased FBN results over the demonstration period.

Below are the steps for determining whether the vendors were budget neutral using the different methodologies. Note that Steps 1 and 2 are the same for all three methodologies. Step 3 differs depending on which of the three methods are used.

Step 1. Calculate demonstration payments for the IG *(Demonstration$_{IG}$):* Payments made to Vendor A and B for providing demonstration services to the IG were provided by CMS' Office of Finance Management (OFM) in a spreadsheet, *complete SRRD Monthly Reports.xlsx,* received on August 2, 2012. Total payments were summed over the three-year demonstration period for Vendor B and a two-year demonstration period for Vendor A, since Vendor A dropped out at the end of Year 2.

Step 2. Calculate total deferred fees (*Total Deferred Fees)*: Total deferred fees were calculated using the total payments made from the OFM spreadsheet (90 percent) and the deferral fee percentage (10 percent):

$$Total\ Deferred\ Fee = \frac{Demonstration_{IG}}{0.90} \times 0.10$$

Step 3: Calculate the per-member-per-month (PMPM) difference in Medicare claims expenditures between the IG and the ACG: The savings (or expenditures) achieved in the IG are calculated using the three different methods:

- **Simple differences** – The per-member-per-month total Medicare expenditures for the ACG are subtracted from those for the IG. The resulting difference represents the savings (or expenditures) per member per month:

  $PMPM_{POST\_IG} - PMPM_{POST\_ACG} = PMPM_{IG\text{-}ACG}$

- **Difference-in-differences** – The difference in the per-member-per-month total Medicare expenditures between the IG and ACG in the pre-demonstration (baseline) period subtracted from the difference in the per-member-per-month total Medicare expenditures between the IG and ACG in the demonstration period. The baseline period consists of January 1, 2008 to April 30, 2009. This method controls for differences in Medicare expenditures in the baseline period before the start of the demonstration.

  $(PMPM_{POST\_IG} - PMPM_{POST\_ACG}) - (PMPM_{PRE\_IG} - PMPM_{PRE\_ACG}) = PMPM_{IG\text{-}ACG}$

- **Regression-based adjustment** – This method estimates the savings controlling for baseline beneficiary-level characteristics. This method controls for differences in demographic characteristics between the IG and ACG groups as well as differences in baseline spending between the two groups. For each vendor, the following regression was run:

  $$P_{i,t=2} = \beta_0 + \beta_1 P_{i,t=1} + \beta_2 age_{i,t=1} + \beta_3 female_i + \beta_4 nonwhite_i + \beta_5 dual_{i,t=1} + \beta_6 IG_i$$

  i: indexes the individual

t: indicates time (*t=1* is the baseline period and *t=2* is the demonstration period)
P: PMPM Medicare expenditures
age: age of beneficiary
female: indicator equal to unity if the beneficiary was female
nonwhite: indicator equal to unity if the beneficiary was not "White non-Hispanic"
dual: indicator equal to unity if the beneficiary was a dual eligible any time during the baseline period
IG: indicator equal to unity if the beneficiary was a member of the IG

The estimated coefficient, $\theta_{6,}$ on IG_i is the difference in IG versus ACG spending for each vendor after controlling for demographic variables and Medicare expenditures in the baseline period.

Note: Because this demonstration is an intent-to-treat study design, once a beneficiary is eligible, he/she is in the calculations throughout the demonstration.[57]

Step 4: Convert the PMPM Medicare savings (or expenditures) from the difference between the IG and ACG into total Medicare savings (or expenditures): The total Medicare savings (or expenditures) achieved for the IG was calculated by multiplying the difference between the per-member-per-month Medicare savings (or expenditures) for the IG and the ACG by the member months (MM) of the IG, using the following formula:

$$Medicare_{IG\text{-}ACG} = PMPM_{IG\text{-}ACG} \times MM_{IG}$$

For Vendor B, the total number of member months (MM_{IG}) is 720,000 (20,000 beneficiaries X 12 months X 3 years); for Vendor A, MM_{IG} is 480,000 (20,000 beneficiaries X 12 months X 2 years).

Step 5: Calculate budget neutrality: Budget neutrality was calculated by summing total Medicare savings (or expenditures), total payments made to the vendor for demonstration services, and the fees that were deferred over the demonstration period, using the following formula:

$$NetCost = (Demonstration_{IG} + Total\ Deferred\ Fees) + (Medicare_{IG\text{-}ACG})$$

Step 6: Reconciliation of deferred fees: Payment of deferred fees was determined as follows:

- If *NetCost < 0*, CMS pays vendor *Total Deferred Fees*
- If *NetCost > Total Deferred Fees*, CMS keeps *Total Deferred Fees*
- If *NetCost < Total Deferred Fees*, CMS pays vendor *Total Deferred Fees – NetCost*

[57] Consistent with the interim budget neutrality analysis, we top coded total Medicare payment at the 99.5 percentile level.

6.3 Findings

Table 3 presents the results of the FBN using Method 1 (simple differences). The demonstration payments made to each vendor are shown in Step 1. The total deferred fees are shown in Step 2. Per-member-per-month Medicare savings or costs calculated using simple differences are shown in Step 3. Step 4 shows the total Medicare savings or costs based on the per-member-per-month Medicare savings or costs. These findings show that Vendor A's IG spent less than its ACG, generating Medicare savings; and Vendor B's IG spent more than its ACG, resulting in Medicare costs. Net costs (Step 5) are then calculated by summing demonstration payments, deferred fees and total Medicare costs or savings (savings would be negative and therefore subtracted). The reconciliation of deferred fees is determined based on net costs (Step 6).

For Vendor A, since net cost is less than zero, based on the reconciliation formula, CMS would pay Vendor A its total deferred fees. For Vendor B, this method yields net costs greater than the amount of total deferred fees so CMS would retain the deferred fees.

Table 3. Summary of Final Budget Neutrality Results Using Simple Differences (Method 1)

	Vendor A	Vendor B
Step 1. Calculate demonstration payments for the IG[a] *Demonstration*$_{IG}$	$1,904,213	$2,384,196
Step 2. Calculate total deferred fees[b] *Total Deferred Fees*	$211,579	$264,911
Step 3. Calculate PMPM difference in Medicare expenditures between IG and ACG using simple differences $PMPM_{IG\text{-}ACG} = PMPM_{POST\_IG} - PMPM_{POST\_ACG}$	-$10.16	$8.51
Step 4. Convert PMPM difference between IG and ACG to total Medicare savings (or expenditures)[c] $Medicare_{IG\text{-}ACG} = PMPM_{IG\text{-}ACG} \times MM_{IG}$	-$4,876,641	$6,125,537
Step 5. Calculate budget neutrality $NetCost = Demonstration_{IG} + Total\ Deferred\ Fees + Medicare_{IG\text{-}ACG}$	-$2,760,848	$8,774,644
Step 6. Reconciliation of deferred fees • If *NetCost < 0*, then CMS will pay vendor *Total Deferred Fees* • If *NetCost > Total Deferred Fees*, then CMS keeps *Total Deferred Fees* • If *NetCost < Total Deferred Fees*, then CMS will pay vendor *Total Deferred Fees – NetCost*	CMS pays *Total Deferred Fees*	CMS keeps *Total Deferred Fees*

[a]Demonstration payments made were provided by OFM in a spreadsheet, *complete SRRD Monthly Reports.xlsx,* on August 2, 2012. Total payments are summed over the three-year demonstration period for Vendor B and two-year demonstration period for Vendor A (Vendor A dropped out at the end of Year 2).

[b]The fee deferral rate of 10 percent was applied to the OFM's demonstration payments to obtain total deferred fees.

[c]For Vendor B the total number of member months (MM_{IG}) is 720,000 (20,000 beneficiaries X 12 months X 3 years) and for Vendor A, MM_{IG} is 480,000 (20,000 beneficiaries X 12 months X 2 years).

Table 4 shows the results of each step when using Method 2 (the DID approach) for comparing the Medicare expenditures of IG to those for the ACG, accounting for differences in the pre-demonstration period IG and ACG groups. Steps 1 and 2 are the same as shown in Table 3.

However, for Step 3, we use the DID approach, according to which both Vendor A's IG and Vendor B's IG Medicare expenditures were less than the Medicare expenditures of their respective ACGs. Multiplying this amount by the number of beneficiaries and the number of demonstration months yields the total Medicare savings (Step 4). Since the resulting net costs for both vendors (Step 5) were less than zero, based on the reconciliation formula, CMS would pay both vendors the total deferred fees (Step 6).

Table 4. Final Budget Neutrality Results Using Difference-in-Differences (Method 2)

	Vendor A	Vendor B
Step 1. Calculate demonstration payments for the IG[a] $Demonstration_{IG}$	$1,904,213	$2,384,196
Step 2. Calculate total deferred fees[b] *Total Deferred Fees*	$211,579	$264,911
Step 3. Calculate PMPM difference in Medicare expenditures between IG and ACG using differences-in-differences $PMPM_{POST\_IG} - PMPM_{POST\_ACG}) - (PMPM_{PRE\_IG} - PMPM_{PRE\_ACG}) = PMPM_{IG-ACG}$	-$9.39	-$4.30
Step 4. Convert PMPM difference between IG and ACG to total Medicare savings (or expenditures)[c] $Medicare_{IG-ACG} = PMPM_{IG-ACG} \times MM_{IG}$	-$6,760,143	-$3,094,553
Step 5. Calculate budget neutrality $NetCost = Demonstration_{IG} + Total\ Deferred\ Fees + Medicare_{IG-ACG}$	-$4,644,350	-$445,446
Step 6. Reconciliation of deferred fees • If *NetCost < 0*, then CMS will pay vendor *Total Deferred Fees* • If *NetCost > Total Deferred Fees*, then CMS keeps *Total Deferred Fees* • If *NetCost < Total Deferred Fees*, then CMS will pay vendor *Total Deferred Fees – NetCost*	CMS pays *Total Deferred Fees*	CMS pays *Total Deferred Fees*

[a]Demonstration payments made, as noted, were provided by OFM in a spreadsheet, *complete SRRD Monthly Reports.xlsx,* on August 2, 2012. Total payments are summed over the three-year demonstration period for Vendor B and two-year demonstration period for Vendor A (Vendor A dropped out at the end of Year 2).

[b]The fee deferral rate of 10 percent was applied to the OFM's demonstration payments to obtain total deferred fees.

[c]For Vendor B the total number of member months (MM_{IG}) is 720,000 (20,000 beneficiaries X 12 months X 3 years) and for Vendor A, MM_{IG} is 480,000 (20,000 beneficiaries X 12 months X 2 years).

Table 5 shows the results of each step when using Method 3 (the regression approach) for comparing the Medicare payments for the IG to those for the ACG while accounting for beneficiary characteristics and differences in the pre-demonstration period IG and ACG expenditures. Steps 1 and 2 are the same as shown in Tables 1 and 2. For Step 3, we use the regression approach which shows that both Vendor A's and Vendor B's IG spent less than their respective ACGs. Multiplying this amount by the number of beneficiaries and the number of demonstration months yields the total Medicare savings (Step 4). For Vendor B, however, the Medicare savings were not enough to cover the demonstration payments made to the vendor. Thus, for Vendor B, CMS would retain the deferred fees based on the reconciliation formula (Step 6). For Vendor A, since the resulting net cost in Step 5 was less than zero, based on the reconciliation formula, CMS would pay the vendor the total deferred fees (Step 6).

Table 5. Final Budget Neutrality Results Using Regression Adjustment (Method 3)

	Vendor A	Vendor B
Step 1. Calculate demonstration payments for the IG[a] *Demonstration*$_{IG}$	$1,904,213	$2,384,196
Step 2. Calculate total deferred fees[b] *Total Deferred Fees*	$211,579	$264,911
Step 3. Calculate PMPM difference in Medicare expenditures between IG and ACG using regression adjustment *PMPM*$_{IG\text{-}ACG}$ =β_6	-$7.39	-$1.62
Step 4. Convert PMPM difference between IG and ACG to total Medicare savings (or expenditures)[c] *Medicare*$_{IG\text{-}ACG}$ = *PMPM*$_{IG\text{-}ACG}$ x *MM*$_{IG}$	-$5,324,172	-$1,164,866
Step 5. Calculate budget neutrality *NetCost = Demonstration*$_{IG}$ *+ Total Deferred Fees + Medicare*$_{IG\text{-}ACG}$	-$3,208,379	$1,484,241
Step 6. Reconciliation of deferred fees • If *NetCost < 0*, then CMS will pay vendor *Total Deferred Fees* • If *NetCost > Total Deferred Fees*, then CMS keeps *Total Deferred Fees* • If *NetCost < Total Deferred Fees*, then CMS will pay vendor *Total Deferred Fees – NetCost*	CMS pays *Total Deferred Fees*	CMS keeps *Total Deferred Fees*

[a]Demonstration payments made, as noted, were provided by OFM in a spreadsheet, *complete SRRD Monthly Reports.xlsx,* on August 2, 2012. Total payments are summed over the three-year demonstration period for Vendor B and two-year demonstration period for Vendor A (Vendor A dropped out at the end of the Year 2).

[b]The fee deferral rate of 10 percent was applied to the OFM's demonstration payments to obtain total deferred fees.

[c]For Vendor B the total number of member months (*MM*$_{IG}$) is 720,000 (20,000 beneficiaries X 12 months X 3 years) and for Vendor A, *MM*$_{IG}$ is 480,000 (20,000 beneficiaries X 12 months X 2 years).

The results shown above use a demonstration period of two years for Vendor A due to its exit at the end of Year 2. Table 6 shows that the resulting reconciliation of deferred fees would remain the same, even if we included Medicare payments incurred in Year 3 for Vendor A. Using a three-year demonstration period, we found that the per-member-per-month costs for the IG were $9.30 lower than those for the ACG. This resulted in a net savings of $4,581,297. The net savings over the three-year period is larger than over the two-year period, since the per-member-per-month savings is multiplied by a higher number of total demonstration months. Table 6 uses the simple differences methodology to illustrate the similarity in results when using a two- versus three- year demonstration period. The same pattern holds when using both the difference-in-differences and regression-based methodologies.

Table 6. Vendor A, Summary of Results by Length of Demonstration Period Using Simple Differences

	2 Year Demonstration Period	3 Year Demonstration Period
$PMPM_{IG\ ACG}$	-$10.16	-$9.30
NetCost	-$2,760,848	-$4,581,297
Reconciliation	CMS pays Total Deferred Fees	CMS pays Total Deferred Fees

6.4 Conclusions and Discussion

The FBN analysis determines whether CMS will pay deferred demonstration fees to the vendors or retain them based on whether each vendor's program was budget neutral or not. Methods 1, 2 and 3 for final budget neutrality reconciliation indicate that Vendor A's program did generate enough savings to cover the demonstration payments made to the vendor and the deferred fees, implying that CMS would pay the deferred fees to Vendor A. Method 2 for the final budget neutrality reconciliation indicates that Vendor B's program did generate enough savings to cover the demonstration payments paid to the vendor and the deferred fees, implying the CMS would pay the deferred fees to Vendor B. However, Methods 1 and 3 indicate that Vendor B did not generate enough savings to cover the demonstration payments made to the vendor, implying that CMS would not pay the deferred fees to Vendor B.

Tables 7 and 8 summarize the results of the alternative methods for each vendor. For Vendor A (Table 7), all three methods resulted in CMS paying the vendor the deferred fees. For Vendor B (Table 8), the difference-in-differences method found that Vendor B's IG did save enough money to cover the payments made to vendors and that CMS would pay the deferred fees to the vendor based on this method. But the simple differences and regression-based methods both found that the IG did not save enough money to cover the payments and that CMS would not pay deferred fees to Vendor B.

Table 7. Vendor A Summary of Results by Method

	Method 1 (Simple Differences)	Method 2 (Differences in Differences)	Method 3 (Regression Adjusted)
PMPM Difference	-$10.16	-$9.39	-$7.39
NetCost	-$2,760,848	-$4,644,350	-$3,208,379
Reconciliation	CMS pays Total Deferred Fees	CMS pays Total Deferred Fees	CMS pays Total Deferred Fees

Table 8. Vendor B Summary of Results by Method

	Method 1 (Simple Differences)	Method 2 (Difference in Differences)	Method 3 (Regression Adjusted)
PMPM Difference	$8.51	-$4.30	-$1.62
NetCost	$8,774,644	-$445,446	$1,484,241
Reconciliation	CMS keeps Total Deferred Fees	CMS pays Total Deferred Fees	CMS keeps Total Deferred Fees

Caution is needed, however in interpreting the budget neutrality results based on Methods 1, 2, and 3. They do not reflect any *causal* financial effects of the SRRD on Medicare expenditures. The FBN methodology is an *accounting* activity that determines whether deferred funds should be retained or distributed to the vendors. For determining the *impact* of the demonstration on Medicare spending, statistical significance is required.

As demonstrated in Table 9 the net cost results were *not* statistically significant at the 1, 5 or 10 percent significance level using Method 1 (simple differences). In other words, IG Medicare spending was *not statistically different* from ACG Medicare spending for either vendor, as can be seen from the confidence intervals resulting from the calculations, which include zero. The lack of statistical significance was also found when using Method 2, the difference-in-differences method and Method 3, the regression-based approach. Thus, we cannot conclude from these findings that IG Medicare expenditures were statistically different from those of the ACG for either vendor.

Table 9. FBN Results Using Simple Differences with 95 Percent Confidence Intervals and P-Values

	Vendor A	Vendor B
Net Cost	-$2,760,848	$8,774,641
95% Confidence Interval		
Lower Bound	-13,594.141	-$6,493,619
Upper Bound	8,072,445	$24,042,906
P Value	0.353[a]	0.401[a]

[a]The p-values of 0.353 for Vendor A and 0.401 for Vendor B are not statistically significant at the 10 percent level.