

Drilling down on dual eligibles

How an analytical
segmentation approach can
help organizations with their
health care quality and
cost-control strategies

About the authors

Leslie Korenda is a research manager with the Deloitte Center for Health Solutions, Deloitte Services LP. She helps Deloitte's health care, life sciences, and government practices with her research to inform health care system stakeholders about emerging trends, challenges, and opportunities. Prior to joining Deloitte, she worked in the private and public sectors in a variety of health care settings, including federal agencies, local health departments, medical centers/health systems, and community health organizations. Korenda received a Bachelor of Science degree from Virginia Tech and a Master of Public Health degree from Yale University.

Daniel Byler, a data scientist with Deloitte Services LP, manages a portfolio of quantitative projects across Deloitte's research agenda. Prior to his current role, he supported clients in large federal agencies on data-focused projects.

Kevin Bingham, ACAS, MAAA, is an Advanced Analytics and Modeling Principal at Deloitte Consulting LLP in Hartford, Connecticut. He is the leader of Deloitte's Claim Predictive Modeling and Medical Professional Liability practices, co-chair of the Casualty Actuarial Society's Innovation Council, past chairperson of the American Academy of Actuaries' Medical Professional Liability Committee, and an official academy spokesperson.

Tory Schiltgen is a consultant in the Advanced Analytics and Modeling practice at Deloitte Consulting LLP. She specializes in applying advanced modeling and information management techniques to reveal valuable insights in critical business areas. With experience in both the health care and banking industries, Schiltgen has developed and supported a variety of innovative predictive modeling business solutions, including lifestyle-based analytics, disease state modeling, health exchange population profiling, medical/claims text analytics, and underwriting application triage modeling.

Gregory Scott, Deloitte Consulting LLP, is the national leader of the Health Plans practice. He has over 25 years of industry experience as a management consultant, insurance company executive, and government official. Greg's experience includes health system reform, Medicare, Medicaid, market strategy, operational performance improvement, and M&A. He has served commercial health plans, integrated delivery systems, PBMs, and federal and state government agencies.

Contents

Drilling down on dual eligibles | 2

How an analytical segmentation approach can help organizations with their health care quality and cost-control strategies

Appendix: Methodology | 12

Endnotes | 13

Contacts | 15

Acknowledgements | 16

About the Deloitte Center for Health Solutions | 16

Deloitte's Advanced Analytics and Modeling (AAM) practice is a market-leading consulting and advanced analytics practice that conducts business analytics advisory work and develops end-to-end custom analytic solutions. AAM maintains a state-of-the-art high-performance data center in Hartford, Connecticut, which is a custom-built facility designed from the ground up for the sole purpose of supporting large-scale data mining activities and research and the development of new techniques and solutions. We employ a team of more than 100 dedicated PhDs, data scientists, and industry subject matter experts to drive business strategy, performance, and decision making for our clients.

Drilling down on dual eligibles

How an analytical segmentation approach can help organizations with their health care quality and cost-control strategies

Analytics: A powerful tool for understanding high-risk populations

There has never been a better time for health care organizations to understand how to efficiently and effectively leverage data analytics to improve care and lower costs. The aging US population, the increased prevalence of chronic disease, and the emergence of payment models that reward value over volume are all driving health care stakeholders to operate more efficiently. At the same time, the tools and know-how to perform analytics, even on very large data sets, have been proliferating. Marrying analytics with the information that health care organizations collect—or could collect—can allow providers and plans to understand their patients and beneficiaries in far more detail than has hitherto been feasible. This, in turn, could create the opportunity to more closely tailor care to the needs of specific patient or beneficiary populations, improving the quality of care while helping to control costs.

The imperative to improve care and lower costs is especially urgent for organizations serving vulnerable, high-risk groups such as “dual eligibles”—individuals who are eligible for both Medicare and Medicaid coverage. Dual eligibles are among the health care system’s most sick, most complex, and most expensive consumers, and research underscores the importance of “establishing well-targeted interventions for specific subsets of the dual-eligible population in order to reduce unnecessary hospitalizations and potentially achieve savings.”¹

This report describes an analytical approach to segmenting the Medicare-eligible population that is substantially more effective in identifying attributes associated with dual-eligible status than a traditional, unsegmented approach. Central to the approach is the use of sociodemographic, lifestyle, and other non-medical data that health plans may collect but do not regularly use to segment or describe patient populations. Health plans could use an expanded version of this approach to target distinct subsets of dual eligibles for care interventions.

This report describes an analytical approach to segmenting the Medicare-eligible population that is substantially more effective in identifying attributes associated with dual-eligible status than a traditional, unsegmented approach.

THE POLICY CONTEXT

While much of the cost of caring for dual eligibles can be attributed to their low income and high health care needs, at least some of the cost is thought to be an artifact of ineffective coordination between Medicare and Medicaid services.² As such, dual eligibles have been the focus of a number of policy initiatives aimed at slowing the growth in this group's health care costs through better service coordination.

The Affordable Care Act of 2010 included several initiatives focused on dual eligibles. One such initiative, the Financial Alignment Initiative (FAI), is testing two new payment models—capitated payments and managed fee-for-service—for the dual-eligible population through a number of state-based “demonstrations.”³ States participating in the demonstrations are responsible for coordinating Medicare and Medicaid benefits and spending for dual-eligible beneficiaries through contracts with private managed-care plans.⁴ Among the United States' 9.6 million dual eligibles,⁵ 1.4 million (15 percent) are getting their care from these demonstrations; of the rest, more than half a million (6 percent) are in non-demonstration managed fee-for-service programs, where the states can share in savings, and the remaining 7.6 million (79 percent) are in traditional fee-for-service Medicare and Medicaid.⁶ More than 70 unique health plans are participating in the FAI across the country.⁷

The value of using expanded data

Our analysis of the peer-reviewed and gray literature, along with conversations with industry experts, suggests that few health plans use sociodemographic and lifestyle data in combination with claims data to segment their dual-eligible beneficiaries. (For two examples of exceptions, see the sidebar “Health plans putting analytics into action.”) This may not be surprising, considering the many challenges that can arise in collecting this type of data. One potential barrier is consumers' concerns about privacy and security: Members may refuse to provide what they may regard as sensitive information about their habits and preferences. Another challenge is that demographic data coming from government agencies or prior health plans or providers may be incomplete or inaccurate when transmitted to a new health plan: Plans often spend a lot of time validating the data received upon enrollment. Finally, one issue specific to dual eligibles is that these individuals, with their low incomes, tend not to generate the shopping pattern and lifestyle analytics information that is being used by many retailers today to understand their customers.

HEALTH PLANS PUTTING ANALYTICS INTO ACTION

Amerigroup, a wholly owned subsidiary within Anthem, is a managed care company serving duals through Medicaid and Medicare Advantage.⁸ As part of its care management solutions, the organization uses social workers and nurses for activities such as nursing home community reintegration and assistance in finding affordable and accessible housing.⁹ To identify, segment, and tailor care coordination programs, the company uses medical and pharmacy claims and demographic data.¹⁰

UnitedHealth has invested heavily in its big data strategy, which includes proprietary analytics and the use of other technologies like health information exchanges. With approximately 700 terabytes of structured and unstructured data, the organization runs queries and models data using a variety of tools, such as business intelligence, data mining, and online analytical processing. Future plans include potentially mining social network data and unstructured data from internal applications.¹¹

Perhaps because of such constraints, many plans take relatively few variables into consideration when segmenting their beneficiaries and running predictive analyses. For instance, many plans tend to rely solely on claims data for predicting which enrollees are likely to be high-risk. The traditional method

of segmenting Medicare beneficiaries is even simpler: Many plans divide them into just two categories, the under-65 disabled and all others. Yet, as we demonstrate here, gathering and analyzing a combination of nontraditional, non-medical data and medical/pharmacy data can allow for a more detailed picture of an individual’s health and risk that can help health plans more clearly understand care needs.

The analysis: Understanding dual eligibles

To demonstrate the potential power of leveraging more information to understand this heterogeneous population, we used data from the Medicare Current Beneficiary Survey (MCBS), an annual nationally representative survey of Medicare beneficiaries administered by the Center for Medicare and Medicaid Services.¹² This survey provides detailed information on health care insurance coverage, spending by payment source (Medicare, Medicaid, private insurance, and out-of-pocket spending), and the use of prescription drugs and long-term care services, as well as information on a rich set of socioeconomic and demographic characteristics, such as marital status, living arrangements, education, and income level. It also provides information on survey participants’ Medicare claims. Our dataset included MCBS records for 15,573 individuals, with 19 percent of the sample identified as dual eligibles.

The analysis proceeded in three stages. First, we conducted a cluster analysis to divide the 15,573 MCBS respondents in our dataset into discrete segments based on similarities and differences in their demographic and health status. Then, we looked at the entire set of 15,573 individuals to determine what attributes, as described in the MCBS data, were significantly associated with dual-eligible status. Finally, we examined each of the resulting segments to determine what attributes were significantly associated with dual-eligible status *within each segment*. The goal was to understand whether looking at each segment independently would allow us to identify attributes associated with dual-eligible status that would have been “hidden” in the analysis of the entire, unsegmented group. The results confirmed that the segment-by-segment analysis did, in fact, identify more attributes connected with dual-eligible status than the analysis of the group as a whole.

Five types of Medicare beneficiaries

The cluster analysis identified five segments of Medicare beneficiaries in the MCBS: “Older and likely to be widowed,” “Young and disabled,” “Older and overweight,” “Older and healthy weight,” and “The survivors.”¹³ To establish the clusters, we used key demographic and health status variables that we believed would paint the most illustrative demographic profile and would produce meaningful clusters. Table 1 shows the variables that were significant in distinguishing the segments.

Table 1. Variables that established the segments

	Older and likely to be widowed (N=3619)	Young and disabled (N=1894)	Older and overweight (N=4620)	Older and healthy weight (N=2507)	The survivors (N=2933)
Body mass index	X		X	X	X
Age	X	X	X	X	X
Depression		X			
Year-over-year change in health status			X		
Marital status	X	X			X

Note: An X indicates that a variable was significant in the cluster—that is, it played a greater-than-chance role in making that group different from the others. In addition to the variables listed in the table, we included education level, gender, income, and smoking status in the model; these characteristics were not significantly different among the segments that we examined.

The segments differed in their sociodemographics, lifestyle variables, health conditions, and health care costs (table 2). Each segment contained roughly

20 percent dual eligibles, except for the “Young and disabled” segment, of whom 71 percent were dual eligibles.

Table 2. Descriptive statistics for each segment

	Older and likely to be widowed (N=3,619)	Young and disabled (N=1,894)	Older and overweight (N=4,620)	Older and healthy weight (N=2,507)	The survivors (N=2,933)	Total (N=15,573)
% of dual eligibles in cluster	21%	71%	17%	20%	21%	19%
Gender						
Female	55.6%	44.3%	50.4%	58.2%	64.7%	54.5%
Male	44.4%	55.7%	49.6%	41.8%	35.3%	45.5%
Age						
Average age	83	43	68	68	85	71
Income						
Less or equal to \$10,000	11.7%	33.6%	9.4%	11.8%	13.0%	12.8%
\$10,001–\$20,000	31.0%	40.7%	20.7%	23.5%	31.4%	26.4%
\$20,001–\$50,000	43.0%	20.1%	47.1%	42.7%	44.2%	42.8%
Greater than \$50,000	14.3%	5.6%	22.8%	22.0%	11.4%	18.0%
BMI						
Underweight	0.0%	2.6%	0.0%	4.8%	9.3%	2.5%
Normal	0.0%	23.6%	0.0%	80.9%	90.7%	31.8%
Overweight	71.7%	31.5%	47.1%	14.3%	0.0%	36.8%
Obese	28.3%	42.3%	52.9%	0.0%	0.0%	28.9%
Negative change in health						
No	73.2%	67.6%	88.6%	71.1%	72.5%	78.0%
Yes	26.8%	32.4%	11.4%	28.9%	27.5%	22.0%
Current or previous smoker						
No	45.6%	42.0%	39.8%	38.5%	53.3%	42.7%
Yes	54.4%	58.0%	60.2%	61.5%	46.7%	57.3%
Depression						
No	81.0%	39.4%	75.0%	75.2%	81.1%	74.1%
Yes	19.0%	60.6%	25.0%	24.8%	18.9%	25.9%
Mental condition count						
0	69.1%	24.6%	71.1%	70.1%	65.6%	66.0%
1	22.3%	35.1%	22.1%	22.2%	22.3%	23.3%
2 or more	8.6%	40.3%	6.8%	7.7%	12.1%	10.7%
Marital status						
Married	43.1%	27.7%	62.7%	56.2%	38.2%	51.3%
Widowed	46.5%	2.2%	15.9%	17.2%	52.7%	25.9%
Divorced/separated	7.8%	28.3%	17.0%	20.0%	5.9%	15.3%
Single	2.6%	41.8%	4.4%	6.6%	3.2%	7.5%

	Older and likely to be widowed (N=3,619)	Young and disabled (N=1,894)	Older and overweight (N=4,620)	Older and healthy weight (N=2,507)	The survivors (N=2,933)	Total (N=15,573)
High school graduation or above						
No	29.3%	28.2%	21.2%	22.2%	28.8%	24.5%
Yes	70.7%	71.8%	78.8%	77.8%	71.2%	75.5%
Physical condition count						
0	3.4%	19.4%	7.1%	11.5%	5.3%	8.1%
1	10.2%	21.6%	16.4%	20.4%	11.0%	15.6%
2	18.0%	21.9%	22.5%	23.7%	18.9%	21.4%
3	19.9%	15.1%	21.2%	17.2%	21.0%	19.6%
4	18.2%	9.6%	14.0%	12.1%	17.6%	14.6%
5 or more	30.3%	12.4%	18.8%	15.1%	26.2%	20.7%
Cost variables						
Part A reimbursement in calendar year (mean)	\$5,216.51	\$3,880.17	\$2,216.35	\$3,088.28	\$6,719.14	\$3,732.64
Part B reimbursement in calendar year (mean)	\$3,817.01	\$4,070.03	\$2,641.12	\$2,892.59	\$3,830.34	\$3,198.11
Total health insurance premiums (mean)	\$1,132.65	\$306.64	\$966.51	\$958.11	\$1,148.42	\$967.41

Dual eligibles: Distinctly different within each segment

To identify attributes significantly associated with being a dual eligible, we used logistic regression on both the entire MCBS dataset and on each of the five segments within the MCBS dataset to identify variables that differentiated between the dual-eligible and non-dual-eligible populations. The results, given in terms of odds ratios, are shown in table 3.

ABOUT ODDS RATIOS

An *odds ratio* is a measure of relationship between an “exposure”—in this case, an attribute such as smoking status, marital status, or diagnosis—and an “outcome”—in this case, being a dual eligible. If the odds ratio is 1, the exposure has no relationship with the outcome; the odds of the outcome occurring given a particular exposure are equal to the odds of the outcome occurring in the absence of that exposure. An odds ratio greater than 1 indicates a *positive* relationship between the exposure and the outcome; that is, the exposure increases the odds of experiencing the outcome. Conversely, an odds ratio of less than 1 indicates a negative relationship between the exposure and the outcome; the exposure *decreases* the odds of experiencing the outcome.

Table 3. Odds ratios for variables associated with dual-eligible status at a statistically significant level

	Total (N=15,573)	Older and likely to be widowed (N =3,619)	Young and disabled (N=1,894)	Older and over- weight (N=4,620)	Older and healthy weight (N=2,507)	The survivors (N=2,933)
Housing variables						
Respondent's home: Apartment			2.61		0.44	
Respondent's home: One-family detached	0.36	0.34	0.50		0.18	0.24
Respondent's home: A two-family or duplex home		0.46	3.77	3.00		
Health/disease variables						
Angina		0.51				
Skin cancer	0.63					
Cancer (non-skin)		0.55				0.44
Osteoporosis			2.32			
Congestive heart failure			3.25			
Has had a stroke			3.28			
History of a psychiatric disorder	1.73		1.82			
Has trouble hearing						2.02
Self-reported health						
Self-reported health is poor compared to others his/her age					3.32	
Smoking status						
Currently smokes					1.89	
Activities of daily living						
Tooth loss makes it difficult to eat		3.26				
Has difficulty eating solid foods because of teeth problems	2.19	4.02	3.38			
Uses special equipment to eat					23.65	
History of military service						
Ever served in the army	0.49	0.38				
Employment variables						
Currently works				0.37		
Education variables						
Less than high school education		5.87		2.49		3.16
Cognitive variables						
Problem making decisions				2.33		
Total condition count						
Total condition count	1.14	1.29	0.83	1.14	1.15	1.20

A look at table 3 yields two important insights. First, many attributes that are associated with the dual-eligible status at the segment level did not show up as being associated with it at the whole-group level. This indicates that the relationships between these attributes and dual-eligible status can be “masked” in analyses that consider the Medicare-eligible population as a whole (that is, without segmentation). Using a broad brush to paint a picture of dual eligibles, in other words, misses potentially clinically significant nuances that emerge when taking a more finely grained approach.

In fact, our complete analysis showed that more than half of the variables associated with dual-eligible status would likely be missed with an unsegmented approach. Overall, we found 71 variables that were significantly associated with dual-eligible status in either the total sample or in the segments. However, 39 of these variables, or 55 percent, were only apparent when we looked within the segments.¹⁴

The second insight is that many of the attributes that are associated with dual-eligible status, especially at the segment level, are attributes that would not typically be assessed on a claims form. Such attributes include not only factors such as an individual’s education level, history of military service, and housing situation, but also factors such as smoking status and difficulty with activities of daily living (ADLs) that would be important in formulating an intervention plan appropriate to the individual. This points out the desirability of capturing and using non-claims data to understand and manage dual eligibles, as claims forms only capture a fraction of the information that plans and providers may find useful.

The odds ratios in table 3 make it possible to develop profiles of the dual-eligible population within each segment of the broader population of Medicare beneficiaries. These profiles, illustrated in figure 1, make it clear that the dual eligibles within each segment look quite different from one another—underscoring the importance of segmentation to

obtain a more accurate and useful view of the dual-eligible population.

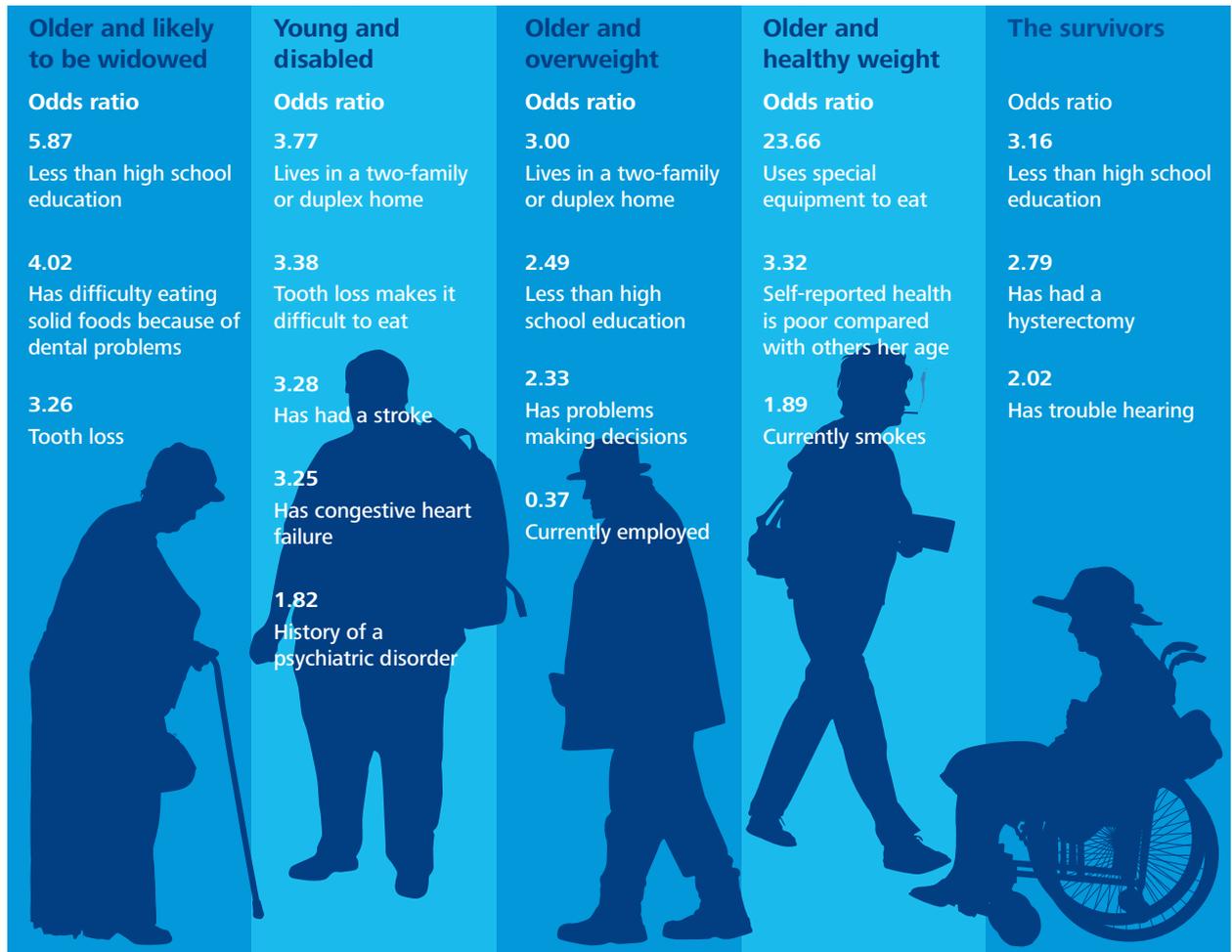
Our analysis may actually understate the level of insight that health plans could derive from an advanced segmentation approach similar to the one outlined above. In order to retain statistical validity, we limited the number of segments to five in order to preserve a minimum sample size of 1,500 people per segment. Health plans, which often have hundreds of thousands or even millions of customers, could develop even more nuanced profiles than the ones that we outline in this study while still identifying segments large enough to have a business impact.

What attributes are most important?

One challenge that accompanies a rich data source is determining the most salient and useful variables or characteristics to consider when trying to understand a population of interest. For health plans, states, and care management companies, this challenge manifests in the problem of how to most efficiently identify vulnerable populations. How can plans determine whom to reach out to in the first 30 or 90 days? What characteristics are most meaningful when trying to separate highest-risk beneficiaries from the rest?

To answer this question for dual eligibles within our five MCBS segments, we conducted a random forest analysis to identify the five variables most strongly associated with dual-eligible status within each segment. The random forest technique accounts for more complex relationships between variables than does logistic regression, allowing it to detect even more nuanced effects that the logistic regressions may not have been able to uncover, and enabling it to rank variables in order of their importance in predicting the outcome of interest (in this case, dual-eligible status).

Figure 1. Profiles of the dual-eligible populations within our Medicare segments



Graphic: Deloitte University Press | DUPress.com

To illustrate the type of insights available from this technique, table 4 gives the top five attributes associated with dual-eligible status in the “older and overweight” segment.

Table 4. Ranked list of variables associated with dual-eligible status

Top five variables associated with dual-eligible status in “older and overweight” segment
Respondent’s home: One-family detached
Wears eyeglasses or contact lenses
Divorced/separated
Current mental illness
Less than high school diploma

The practical value of performing a random forest analysis is that it can allow health plans to understand what to look for first when identifying populations for potential care interventions. In practice, it can be expensive and time-consuming to gather new information about a given patient population, making it important to identify variables that have the greatest fundamental strength over a wide set of population subsets. Random forests (a cousin of decision trees), fill this gap by performing two key functions: splitting the population into deep segments and explicitly ranking variable worth. Due to the way a random forest works, it avoids simply looking for variables that have the greatest strength when considered in isolation; it helps search for

variables that, when in the presence of other variables, may have significant predictive value. By taking such a detailed approach, plans can have increased confidence that the variables they choose to examine will be valuable, not only in an original study of the population of interest, but also in future attempts to combine their newly collected variables in insightful ways.

How the analysis links back

Our analysis points to two main implications. The first concerns the value of using expanded data to understand patient and beneficiary populations. The analysis described above shows that using psychosocial, lifestyle, behavioral, and other non-medical variables are useful both in segmenting broad populations of beneficiaries and in identifying the distinguishing attributes of certain groups—in this case, dual eligibles—within those segments. For example, wearing eyeglasses, a history of military service, and employment status, along with many other characteristics, are not likely to be found in claims data. Yet we found these attributes to be significantly associated with certain clusters in the Medicare population and associated with dual-eligible status.

Plans or health systems could use non-claims data to supplement claims data, forming the basis for the tailoring of care. Such assessments would need to be collected regularly to be accurate and actionable,¹⁵ and they may run into challenges due to privacy and security concerns or other complicating factors. However, these challenges are not insuperable. In fact, one could argue that lifestyle, demographic, and other nontraditional data could be *easier* to come by for dual eligibles than for other groups of beneficiaries. This is because many FAI demonstrations, as well as non-FAI

programs serving dual eligibles, collect detailed data on beneficiaries in order to perform a health risk assessment (see sidebar, “Targeting care through health risk assessments”).

The second implication is that, by using an analytical approach to understanding their dual-eligible beneficiaries, health care organizations can tease out interventions to manage this heterogeneous population. An analytical approach such as the one demonstrated here can offer insights and a method to stratify members for different interventions. Because each segment has unique characteristics associated with being a dual eligible, health care organizations can identify which characteristics are most important to better serve the individuals within the segment. For example, for those segments where social support may be lacking (for instance, those living alone or widowed), that lack of social support may help predict an admission to a nursing home or an acute admission. Some segments are at higher risk for multiple comorbidities. And for healthier segments, interventions that focus on wellness may be just as effective as higher-cost interventions.

In summary: Segmenting the beneficiary population to develop well-targeted, tailored interventions is more likely to yield positive results, in terms of both health and cost outcomes, than taking a one-size-fits-all approach. Analytical methods such as those described here can help organizations effectively segment their Medicare beneficiaries to gain a more finely grained understanding of various sub-populations of dual eligibles. Many organizations have the data to adopt and implement the model we have used. Successfully building and deploying the model and using the information to target and coordinate care could make a meaningful difference in providing better care to this diverse population at a lower cost.

TARGETING CARE THROUGH HEALTH RISK ASSESSMENTS

California: Cal MediConnect

California's FAI demonstration, the Coordinated Care Initiative (CCI), delivers a full continuum of services to dual eligibles through the capitated payment model. Beneficiaries participating in CCI can access acute, primary, institutional, and home- and community-based services (HCBS). Participating health plans are using health risk assessments (HRAs) to identify beneficiaries' needs around medical care, behavioral health, chronic conditions, activities of daily living, and other areas. Cal MediConnect plans perform HRAs on high-risk beneficiaries* within 45 days and on low-risk beneficiaries within 90 days of enrollment, identifying high-risk enrollees using a beneficiary's historic Medicaid and Medicare fee-for-service utilization data.¹⁶ As of September 2014, 78 percent of Cal MediConnect HRAs were completed within 90 days of enrollment.¹⁷

Massachusetts: One Care

Three health plans are participating in Massachusetts' FAI demonstration, One Care, which launched in late 2013. Interdisciplinary care teams—which include the beneficiary, a care coordinator, and an independent-living and long-term services and supports coordinator—work together to develop a customized care plan reflective of each beneficiary's needs and preferences. Approximately 95,000 individuals are eligible for the program, and more than 17,000 were enrolled as of October 2014.¹⁸

Within the first 90 days of a beneficiary's enrollment, One Care plans must conduct a comprehensive assessment of each individual. Beneficiaries may choose the setting in which they prefer to complete the assessment, and family members and other providers can be present. The assessment tool identifies immediate needs and current services, as well as the individual's health conditions, medications, functional status, and more. This assessment is then turned into the core components of an individualized care plan, which includes steps that the beneficiary and One Care team will take to address goals and concerns.¹⁹ The assessment is also used to assign beneficiaries to four rating categories.²⁰

Arizona: Arizona Health Care Cost Containment System

Arizona pulled out of the FAI demonstration program in early 2013, citing slow implementation progress and uncertainty around the future of the program as chief complaints. However, the state will continue to test innovative models of care for its dual eligibles through the Arizona Health Care Cost Containment System (AHCCCS),²¹ which has provided care to its Medicaid and dual-eligible population through contracts with managed care organizations for more than 30 years.²² All plans in the AHCCCS program use health risk assessments and predictive modeling to target appropriate interventions.

** High-risk beneficiaries are defined as those who are "at increased risk of having adverse health outcomes, or worsening of their health and functional status, or whose health conditions require careful monitoring and coordination of multiple medical, long-term services and supports, or behavioral health services."²³*

Appendix: Methodology

Segmentation analysis

In order to group our 15,573 records into mutually exclusive groups with similar characteristics, we used the unsupervised K-means clustering analysis technique.²⁴ The iterative algorithm leveraged demographic and health status variables to statistically separate the data into five unique clusters (table 1). Health care spending and utilization data were not used to define or establish the segments, but were used to describe the segments (that is, how cost differs among the segments). Not all variables were significantly different among the populations; while they were included in the model, they did not have a strong impact on the clusters.

Logistic regression using cross-validation

In performing our regression analysis on the five different clusters, we chose to divide each cluster into five random but equal parts. This allowed us to run the model five different times for each cluster, using 80 percent of the data to train the model and 20 percent of the data to validate the model. For example, the first iteration of the model used data parts 1, 2, 3, 4 of each cluster to build the model, and data part 5 of each cluster to validate the results. The second iteration of the model used data parts 1, 2, 3, 5 of each cluster to build the model and data part 4 of each cluster to validate the results. This process was repeated until

we generated five sets of model coefficients for each cluster. The final model coefficients for each cluster were generated by taking the average of the coefficients across each predictive variable (for example, age, body mass index, marital status, ethnicity, and so on).

Some variables had a high rate of correlation to each other. To deal with these, we used the first principal component of these groups of variables as a way to reduce this problem. For the most part, these did not turn out to be key variables, but they did serve to inform our model selection.

Random forest

In conjunction with performing a logistic regression on each cluster, we also applied an ensemble method of classification known as a random forest to each cluster. To be consistent with our logistic regression analysis, the exact same division of each cluster into five equal parts was used in fitting the random forests. In other words, the data used in the logistic regression analysis was identical to those of the random forest analysis. However, due to the robust nature of the random forest, the model was only fit once to each cluster using a standard 80/20 split within each cluster for training and validation data, respectively. For each cluster, a random forest was fit using a maximum number of trees ranging anywhere from 50 to 500. Each tree within a random forest was trained using a 60/40 split of the training data for each cluster.

Endnotes

1. R. Brown and D.R. Mann, *Best bets for reducing Medicare costs for dual eligible beneficiaries: Assessing the evidence*, Mathematica Policy Research, commissioned by Kaiser Family Foundation, October 2012.
2. Center for Medicare and Medicaid Services, “Fact sheet: People enrolled in Medicare and Medicaid,” March 2015, https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_Factsheet.pdf.
3. CMS.gov, “State proposals,” <http://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/FinancialAlignmentInitiative/StateProposals.html>.
4. CMS.gov, “CMS Financial Alignment Initiative,” <https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/FinancialAlignmentInitiative.html>, accessed September 25, 2015.
5. Medicare Payment Advisory Commission (MEDPAC) and the Medicaid and CHIP Payment and Access Commission (MACPAC), “January 2015 Data Book: Beneficiaries dually eligible for Medicare and Medicaid,” January 9, 2015, <http://medpac.gov/documents/data-book/january-2015-medpac-and-macpac-data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid.pdf?sfvrsn=2>), accessed April 8, 2015.
6. 2014 Aetna Investor Conference, December 11, 2014, <http://www.aetna.com/investors-aetna/assets/documents/2014-investor-conference/2014InvestorConferencePresentation.pdf>, accessed April 4, 2015.
7. Deloitte analysis of health plans participating in the Financial Alignment Initiative.
8. Amerigroup RealSolutions in Healthcare, “Dual eligibles fast facts,” <http://www.amerigroup.com/sites/amerigroup.com/files/files/Dual-Eligibles.pdf>, accessed April 4, 2015.
9. H. Meyer, “The coming experiments in integrating and coordinating care for ‘dual eligibles,’” *Health Affairs*, June 2012, 31(6): 1151–55, <http://content.healthaffairs.org/content/31/6/1151.full?ijkey=VzBe5Hk6ZrkvQ&keytype=ref&siteid=healthaff>, accessed April 4, 2015.
10. Amerigroup RealSolutions in Healthcare, “Programs and practices, external communications, Amerigroup case management and coordinated care best practices,” <http://www.amerigroup.com/sites/amerigroup.com/files/files/Coordinated-Care.pdf>, accessed April 4, 2015.
11. B. Eastwood, “Embracing big data helps UnitedHealth scale, innovate: Insurer applies analytics to many areas of the business,” *FierceHealthIT*, April 6, 2015, http://www.fiercehealthit.com/story/embracing-big-data-helps-unitedhealth-scale-innovate/2015-04-06?utm_medium=rss&utm_source=rss&utm_campaign=rss, accessed April 8, 2015.
12. CMS.gov, “Medicare Current Beneficiary Survey (MCBS),” <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MCBS.html>.
13. The cluster analysis was limited to the identification of five clusters due to considerations of sample size: With only 15,573 data points, identifying a larger number of clusters would have resulted in too few individuals per cluster to conduct the logistic regression models and random forest analyses. Health plans working with larger data sets would, in theory, be able to reliably identify more clusters for a still more finely grained analysis.

14. Note that table 3 does not give the complete list of 71 variables, but presents a selection of the variables where the effects were most clear. One reason we chose not to show all 71 variables is that some of these were “decomposed” out of more complex variables; for example, multiple variables were generated based on multi-level categorical variables like education level (less than high school, college-educated, and so on), while other variables (such as income) had multiple break points to account for the fact that income effects can be non-linear. Additionally, some of the variables exhibited extremely high effect values, which were likely the result of sparse populations. However, even in the results presented in table 3, our overall finding remains the same: The segments show significant variation that is not reflected in the analysis of the overall sample.
15. Center for Health Care Strategies, “Health plan approaches to locating new Medicare-Medicaid enrollees and building trusting relationships,” webinar, January 13, 2015, <http://www.chcs.org/media/CHCS-PRIDE-Webinar-Locating-Medicare-Medicaid-Enrollees-FINAL.pdf>, accessed April 4, 2015.
16. Calduals, “Cal MediConnect health risk assessment (HRA) data now available,” <http://www.calduals.org/2015/02/11/cal-medicconnect-health-risk-assessment-hra-data-now-available/>.
17. California Department of Health Care Services, Cal MediConnect health risk assessment (HRA) dashboard,” http://www.calduals.org/wp-content/uploads/2015/02/HRA_Dashboard-Apr-Sept2014.pdf.
18. Massachusetts Medicaid Policy Institute, “One Care enrollment,” http://www.massmedicaid.org/sites/default/files/download/publication/OneCare_Infographic_Final.pdf
19. Massachusetts Medicaid Policy Institute, “One Care initiative for non-elderly dual eligibles,” http://www.massmedicaid.org/sites/default/files/download/publication/OneCare_Report_Nov2014_Final.pdf.
20. Commonwealth of Massachusetts Executive Office of Health and Human Services and United States Department of Health and Human Services, “Request for responses from integrated care organizations,” <http://www.mass.gov/eohhs/docs/eohhs/healthcare-reform/120730-duals-demo-rfr.pdf>.
21. AHCCCS, “Arizona capitated financial alignment demonstration withdrawal,” <http://www.azahcccs.gov/reporting/Downloads/Integration/ArizonaCapitationFinancial-AlignmentDemonstrationWithdrawal.pdf>.
22. Health Affairs, “Home page,” <http://content.healthaffairs.org/content/16/4/194.full.pdf>.
23. California Department of Health Care Services, “Health risk assessments in the Cal MediConnect,” <http://www.calduals.org/wp-content/uploads/2015/08/Health-Risk-Assessment-Explained.pdf>.
24. Wikipedia, “k-means clustering,” http://en.wikipedia.org/wiki/K-means_clustering.

Contacts

Kevin M. Bingham, ACAS, MAAA

Principal

Deloitte Consulting LLP

+1 860 725 3056

kbingham@deloitte.com

Yvon H Pho, PhD, PMP

Senior manager

Deloitte & Touche LLP

+1 202 368 5644

vpho@deloitte.com

Gregory Scott

Vice chairman, Health Plans national sector
leader

Principal

Deloitte LLP

+1 412 402 5051

grescott@deloitte.com

Harry Greenspun, MD

Director

Deloitte Center for Health Solutions

Deloitte Services LP

+1 202 370 2284

hgreenspun@deloitte.com

Sarah Thomas, MS

Director of research

Deloitte Center for Health Solutions

Deloitte Services LP

+1 202 220 2749

sarthomas@deloitte.com

Acknowledgements

The authors would like to thank the following professionals for their contributions to this article:

A special thanks to **Divya Sharma** and **Yvon Pho** for their work on the analysis and interpretation of the results in this paper. We would like to thank **Mike Petrillo** and the Advanced Analytics and Modeling Data Center along with **Abhishek Kuppili** for their help on this project.

For their work on the concept and writing of this report, we would like to thank **Sarah Thomas**, **Claire Cruse**, and **Elizabeth Stanley**, without whom this report would not have been possible.

We wish to thank **Junko Kaji**, **Bill Copeland**, **Megan Cormier**, **Lucia Giudice**, **Kevin Kenny**, **Jim Whisler**, **Mark Bethke**, **Russ Pederson**, and the many others who contributed their ideas and insights to this project.

About the Deloitte Center for Health Solutions

The Deloitte Center for Health Solutions (DCHS) is the research division of Deloitte LLP's Life Sciences and Health Care practice. The goal of DCHS is to inform stakeholders across the health care system about emerging trends, challenges, and opportunities. Using primary research and rigorous analysis, and providing unique perspectives, DCHS seeks to be a trusted source for relevant, timely, and reliable insights.





Follow @DU_Press

Sign up for Deloitte University Press updates at DUPress.com.

About Deloitte University Press

Deloitte University Press publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte University Press is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Copyright © 2015 Deloitte Development LLC. All rights reserved.
Member of Deloitte Touche Tohmatsu Limited