Engagement Scoring for Care-gap Intervention Optimization

MohamadAli Torkamani, Malhar Jhaveri, Jynelle Mellen
Michael Brown-Hayes, James Chung, Penny Pan, Hakan Kardes
{FirstName}.{LastName}@CambiaHealth.com
Cambia Health Solutions
Portland, OR, 97201

ABSTRACT

Timely preventative health screenings can be crucial for the early detection of serious disease or complications from chronic conditions, such as diabetes. Influencing the right people to obtain recommended screenings, at the right time, can result in significantly improved health outcomes. These screenings, if not attained, are called care-gaps, while performing the screening is called a closed care-gap. The spectrum of individuals managing their own health care ranges from minimal engagement (no screenings in this case) to high engagement (timely screenings). For those who do not obtain timely screenings, we have identified two types: individuals who will respond to outreach and those who will not. Therefore, our focus becomes identifying the right people (those who need to close a gap and have a likelihood of responding) at the right time. This approach will maximize the effectiveness and impact of outreaches. Our recommendation model generates a ranking order where the individuals who are most likely to close their care-gaps after intervention, are ranked first. Our method shows successful results in detecting patients who need a prompt, and our experimental results show that by using this recommendation model, we can increase the number of closed gaps.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; **Health care information systems**; • **Information systems** → **Recommender systems**;

KEYWORDS

Recommendation Impact, Care-gap Closure, Right Time Intervention

1 GAPS IN CARE

Based on condition, age, and sex, the US Preventive Services Task Force (USPSTF) and Centers for Disease Control and Prevention (CDC) publish guidelines for how Americans can best manage their

health by performing needed preventive screenings. Care-gaps are the result of obstacles preventing patients and physicians from implementing care recommendations. Some barriers include misunderstanding of guidelines, lack of awareness, lack of proper transportation to clinics and hospital, fear of procedures like colonoscopy, etc.

For breast cancer screenings, the USPSTF recommends that women 50-74 years of age receive one mammogram every 27 months. For colorectal cancer screenings, the USPSTF recommends that individuals 45-75 years of age receive either one fecal occult blood test (FOBT) every year, one flexible sigmoidoscopy every five years, or one colonoscopy every ten years [2, 6].

There are also specific guidelines for people with diabetes to help manage their care. The CDC encourages individuals with diabetes to annually receive at least one hemoglobin A1c (HbA1c) test to understand their average blood glucose levels, at least one dilated eye exam for early detection of retinal changes, and at least one nephropathy test to check kidney function [5].

The National Committee for Quality Assurance (NCQA) sets guidelines to evaluates the performance of every health plan based on those guidelines [1]. The Healthcare Effectiveness Data and Information Set (HEDIS) measures the performance of health plans based in part on care-gap closure rates of Breast Cancer Screening (BCS), Colorectal Cancer Screening (COL), and Comprehensive Diabetes Care (DIAB).

To improve HEDIS performance, health plans employ clinical staff to develop intervention plans for contacting patients with outreach intended to encourage them to close their care-gaps. Supporting this intervention are algorithms designed to identify members who should have received at least one of these screenings but have not done so.

The number of people covered by a health plan who are also eligible for the above-mentioned screenings can be quite large, and comprehensive outreach campaigns could take several months to complete. During such a campaign, everyone who is eligible for screening will receive a telephone call (except those who have opted-out on the "Do Not Call" list). If several outreach attempts are unable to contact a person, then a letter outlining the relevant information is mailed to their address.

In general, some portion of the population is not *engaged* in their own health. They may be resistant to interventions and may ignore phone calls and letters. This sub-population is less likely to close their care-gaps even if they are contacted several times via various channels. Some people might request that their names be added to the Do Not Call list. On the other hand, a complementary portion of the population are highly *engaged* in their health. They do not require interventions at all. They will close their care-gaps on their own. Finally, there is a group of people who will likely close their

care-gaps, but only after being contacted. One of the challenges of the healthcare system is to prioritize resource allocation to the patients who are higher risk and are more likely to be impacted positively.

To address this problem, we have designed a machine learning recommender system that ranks the more impactable care-gaps higher – i.e., the ones that are more likely to be closed âĂŞ for each person. Some of the screenings such as colonoscopy and eye exam for diabetes are harder to close. Our model can be used to frame a personalized message for each member based on their *engagement score*, and the degree of difficulty in closing their open gaps. If a person is likely to close at least one care-gap, the system recommends this person to the care advocates for intervention.

2 MANIFEST SPACE

To create a reliable feature vector, we have used both direct feature extraction from the data and a collaborative filtering method for data imputation. The data we observe in the claims-record we term manifests.

2.1 Data

We have used four **manifests** categories of data sources to create our features: pharmaceutical claims, specialties of the providers that patients have been visiting, diagnoses made, and the final services performed.

Data used for creating the model is derived from claims data with approximately 30 million rows and over 3 million individuals, containing information such as their gender (male, female and unknown), dates of birth and death (if applicable). Based on a patients age, we categorized people in twelve age groups (≤1, 1-4, 5-12, 13-17, 18-25, 26-35, 36-45, 46-55, 56-65, 6675, 76-85 and ≥86). The data is aggregated to the industry standard quarters from year 2012 to 2017.

The features include everything that a member could claim from a payer (medical rehab, surgeries, treatment for health disorders, prescriptions, etc). within the applicable period. Also, in the data set, each person might be present in multiple year/quarters.

To create an extended feature vector, we constructed a bag-of-words representation for the presence of every possible value that a manifest could have. For example, we used several therapeutic groupers for pharmaceutical data. And, we used the number of times that the patient had a prescription for a specific medication in one calendar quarter as the corresponding feature for that drug. We used the same count-based representation for all other features.

The feature vector was also augmented with patientsâĂŹ demographics in the calendar quarter of interest. This included their age, gender, and several features from the United States census data based on their home neighborhood.

We hand-crafted some feature that we expect to indicate health engagement. In particular, we constructed several features for their medication-adherence based on how timely they are in refilling their recommended prescriptions.

2.2 Smoothing and Missing Value Imputation

A problem with claim data is that missing values do not necessarily mean that the patients have not had a manifest. Besides the noise and human error, the missing values could be caused by the complicated structure of the healthcare system in the United States. For example, a value not being present in a patient's manifest could be due to their multiple coverages or lack of eligibility for specific periods of time. To deal with this problem, we assume that similar patients require similar types of care. Also, many of the features co-occur. For example, many of the diagnoses are comorbidities that patients have at the same time, or certain drugs are always prescribed for specific ailments. As a result, the table of our features for all the patients at all time should form a low-rank matrix. We use a low-rank matrix completion approach for both filling the missing values as well as smoothing the features and removing the noise [4].

Our approach is similar to the Robust PCA method by Candès et. al. [3]. Let X_{ij} be the observed value for feature j for the ith observation (i.e.,p̃atient, year and quarter). We learn a low-dimensional approximation M of the full matrix X, by solving the following nuclear norm minimization convex program:

$$\underset{M}{\text{minimize}} \quad \|M\|_* + \lambda \sum_{ij} |X_{ij} - M_{ij}|$$
$$\text{subject to} \quad \sum_{ij} (X_{ij} - M_{ij})^2 \le \epsilon$$

$\|M\|_*$ is the nuclear norm of the smoothed data, which encourages M to be low-rank. $\sum_{ij} |X_{ij} - M_{ij}|$ is the ℓ_1 norm of the distance between the observations and their approximations and allows the existence of some outlier vaues, while it pushes matrix M to be close to observed values of matrix X as much as possible. From a statistical point of view, this term assumes that there is a sparse set of outliers that can be modeled using a Laplacian distribution. The constraint $\sum_{ij} (X_{ij} - M_{ij})^2 \le \epsilon$ encourages the closeness of M and X in general, but in specific it limits the effects of Gaussian additive noise such as variations in the number of prescription of the same medication for similar people by different doctors. The hyperparameter λ and ϵ are tuned by cross-validation using held-out samples after model training explained in the next section.

3 ENGAGEMENT MODEL

We designed an ensemble model that predicts the likelihood of closing care gaps after a phone call. For the people who are more likely to close their care-gap after a call, we recommend a higher priority for outreach, because the data and model show that we can impact them.

3.1 Predictive modeling of care-gap closure using experimental data

Members with open care-gaps can be grouped into three groups.

(1) Members who will close their care-gaps by themselves.
(2) Those who will respond to an outreach by closing their care-gaps.
(3) Those who are not engaged with their healthcare and will not close their care-gaps, even after outreach. (Figure 1)

With the experimental dataset described in the protocol below, we will be able to estimate the following at an individual-member level.
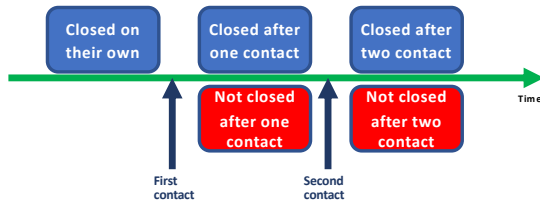
Figure 1: Segmentation of the population based on their behavior in closing their care gaps before being contacted, or after one and two interventions.

(a) Probability of closing care-gaps without outreach.
(b) Using 1 and 2, we also estimated the increased likelihood of closing care gaps after outreach. In other words, we will be able to measure the value it adds to contact a person, and how the likelihood of care gap closure increases accordingly.
(c) Probability of not closing care-gaps with outreach.

3.2 Implementation of the Score

The input to our model is the smoothed and imputed feature vectors, as well as gold standard targets from intervention in the previous years. After feature imputation, we use a hybrid ensemble method consisting of random forest and a support vector regression model for computing the probability of being impacted by an intervention.

The output scores from the model were used to prioritize the member list for closing care-gaps for Medicare and Commercial lines of business.

Experimental Study design was a randomized controlled study of Medicare members with an open care-gap for at least one measure from annual wellness visit, colorectal cancer screening, and breast cancer screening. Randomization was done at a member-level using stratification by engagement model score, i.e. samples were randomly selected from each decile of the score distribution. Below is the study design diagram (Figure 2). The interventions were performed for 10,045 unique members.

Figure 2: Randomized study design – *Engagement score is the output from Data Science Model that will predict who will likely to close the gaps (higher score is better)
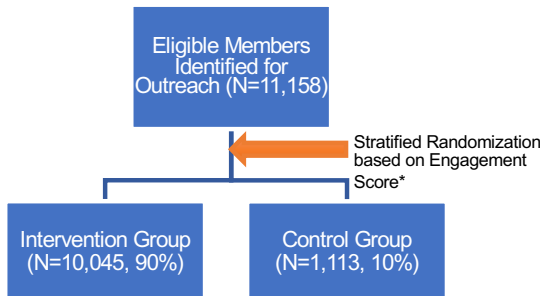
4 EXPERIMENTAL RESULTS

While the model can identify the likelihood of closing care-gaps, we are unable to calculate the likelihood of impacting care-gap closures

Figure 3: A: More Likely To Close by themselves or after a small nudge, B: Likely to Close after intervention, C: Calling them does not change their behaviour, and the are not likely to close their gaps. D: Score-Based Prioritization E: Heuristic (Random) Prioritization. Note that prioritization of the top 100 consumers will gain 288% improvement using the proposed model.
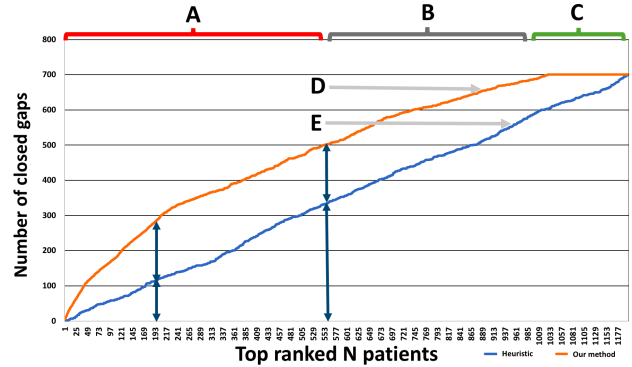
Medicare Measure	Closed after intervention		Not closed after intervention		(p-value) *: statistically significant
	N	Avg. score	N	Avg. score	
Wellness	1,184	**0.58**	3,783	**0.5**	**<0.01***
COL	797	**0.6**	4,572	**0.52**	**<0.01***
BCS	461	**0.58**	1,716	**0.51**	**<0.01***

Table 1: Effectiveness of the engagement model for the Medicare line of business

Medicare Measure	Closed after intervention		Not closed after intervention		(p-value) *: statistically significant
	N	Avg. score	N	Avg. score	
Wellness	228	**0.59**	885	**0.51**	**<0.01***
COL	140	**0.58**	973	**0.52**	**<0.01***
BCS	163	**0.58**	524	**0.52**	**<0.01***

Table 2: Effectiveness of the engagement model for the control group only. Note that while there have been no interventions for control group, the time interval of waiting for care-gaps to be closed has been consistently selected for both the intervention and the control groups.

with outreach, directly from data. To jointly measure the performance of our model as well as the effectiveness of interventions, an experimental study was designed.

Out of 10,045 members, 9,768 members could be contacted. Distribution of measures for outreached members is shown in Figure 4 (highest for colorectal screening following by Wellness and breast cancer screenings.

Table 1 presents the effectiveness of the engagement model. Here the effectiveness is measures by comparing the average engagement

Medicare Measure	Closed after intervention		Not closed after intervention		(p-value) *: statistically significant
	N	Avg. score	N	Avg. score	
Wellness	956	**0.57**	2,898	**0.50**	**<0.01***
COL	657	**0.60**	3,599	**0.52**	**<0.01***
BCS	298	**0.57**	1,192	**0.51**	**<0.01***

Table 3: Effectiveness of the engagement model for the intervention group only

score between members who closed the gaps versus who didnâĂŹt closed the gaps for the respective eligible measures. As you can see for all the three measures, members who closed the respective gaps had a significantly higher engagement score.

In Table 1, we show the effectiveness of the model within the whole population, i.e., both the control and intervention groups combined. To study the performance of the model itself, we should also investigate how the members of the control group behaved regarding their open care-gaps in the period following the intervention campaign. To do so, we performed analyses similar to the process for Table 1, and we analyzed the intervention and control groups in tables 2 and 4 separately. As the results in Table 2 state, the model has been able to successfully identify the people who are engaged in their health and have closed their care-gaps without being contacted during this campaign. Table 4 is also aligned with tables 1 and 2, and it shows that the model has performed similarly well for the intervention group.

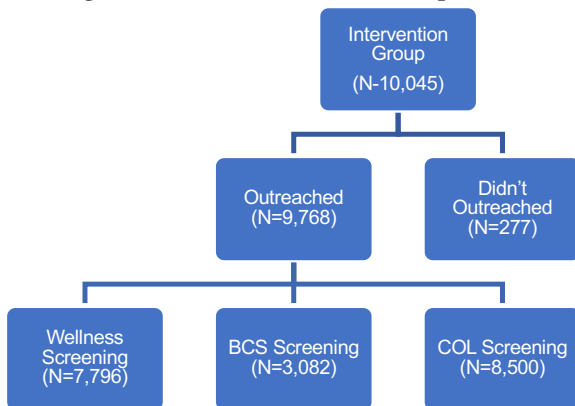**Figure 4: Intervention Outcome Population**

Table 4 shows the effectiveness of interventions in closing gaps for three measures. There is a statistically significant positive lift for wellness and colorectal measures (absolute difference of +4.3% and +2.8%, and relative lift of +20.97% and 22.22%). Lift is the gap-closure percentage difference between intervention and control groups. Lift for breast cancer screening measure is negative but not statistically significant. Negative lift is partly due gender specific measure and during the randomization process members in control group were selected at a member level and not at a measure level.

Also, the distribution of the engagement score for members who closed versus who didnâĂŹt closed the wellness gaps is shows in Figure 5. The patients who closed wellness measure had significantly higher engagement score. Similar patterns were observed for Breast and colorectal cancer screening measures (Figure 5). This verifies that our reommender system has successfully selected the impactable people for outreach.

Figure 5: The distribution of scores of members who closed their wellness gaps vs. the members who didn't close. We observe a significant difference between the two cohorts' predicted scores.
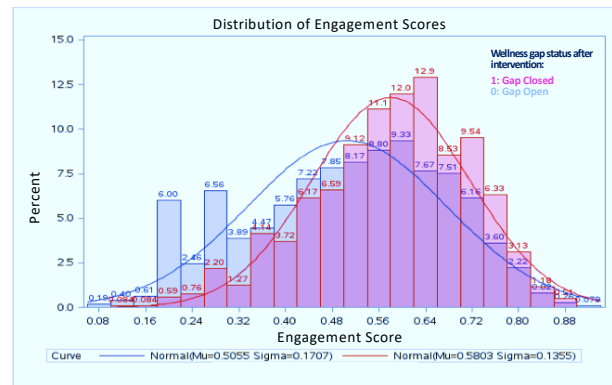
5 DATA ETHICS

In building our data science models and algorithms, we evaluate and test for accuracy, precision and fairness to account for potential biases. We also ensure that the privacy of individuals are respected. It is critically important to build models that are primarily beneficial for our customers. Both through traditional and digital customer interfaces, these models should also create, deepen and maintain consumers' trust.

One of the data ethics considerations within the care gaps model included the control group size and experimental size. Specifically, while the typical ratio would be 30% to 50% or a control group, we opted for 10%. This 10% was also randomly chosen from stratified scores to confidently avoid an accidental accumulation of specific population in the control group. This way, our efforts would more rapidly include as many individuals as possible, but also provided enough data for statistically significant conclusions to assess the effectiveness of interventions and the model in one study. It is important to note that our model does not exclude individuals from being contacted. Instead, it prioritizes based on impact-ability at the right time. Individuals in the control group were not excluded from other campaigns outside the study period.

Other ethical considerations were related to how we apply the scoring within the model. For example, part of the scoring includes the number of gaps an individual has and the severity. We prioritize people with more open gaps higher, regardless of their engagement scores. To best understand these details and how to translate the details into a meaningful and ethical model, we have closely collaborated with key subject matter experts within the organization.

Medicare Measure	Total number of subjects in the study	Intervention Group			Control Group			Lift% (p-value) *: statistically significant
		$N_{contacted}$	N_{closed}	closed (%)	$N_{control}$	N_{closed}	closed (%)	
Wellness Measure	4,967	3,854	956	24.80%	1,113	228	20.50%	Abs.: +4.3% (<0.01*) Rel.:+20.97%
COL Measure	5,369	4,256	657	15.40%	1,113	140	12.60%	Abs.:+2.8% (<0.01*) Rel.:22.22%
BCS Measure	2,177	1,490	298	20.00%	687	163	23.70%	Abs.:-3.70% (>0.1) Rel.:-15.61%

Table 4: Effectiveness of interventions: We observe that interventions improve care-gap closure for wellness and colorectal measures by 4.3% and 2.8% with statistical significance. However, the impact of interventions for breast cancer screening is negative, although the difference is not statistically significant. (Abs.: absolute change, Rel.: absolute change normalized by control baseline.)

6 CONCLUSION

Care gaps closure in not only financially important for the healthcare system, but also it directly helps patients' well-being by identifying conditions at early stages.

Our proposed recommendation system generates ordered prioritization of the patients who are more likely impacted by phone interventions. The results show that in practice, people with high engagement scores are more likely to close their care-gaps after being outreached. Our recommender system could prioritize the outreach and can differentiate who is likely to close the gaps after an intervention.

This model can be extended in several directions. For example, if we afford to contact many people, we can use the system for personalized messages during the outreach. People who are not much engaged in their healthcare, might be motivated by more incentives and encouragement, but highly-engaged people might be willing to have more information about other health-related activities.

We can also use this system, to contact people who need more time to close their care-gaps first and reach to people who will respond faster afterward.

It is easy to add other measures to this system. The model is built on healthcare industry standards (e.g., ICD-10, CPT codes, therapeutic classes). Therefore, it can be used by a broader population. The engagement score can also be used as a proxy for general healthcare engagement for marketing applications.

REFERENCES

[1] Nancy D Beaulieu and Arnold M Epstein. 2002. National Committee on Quality Assurance health-plan accreditation: predictors, correlates of performance, and market impact. *Medical Care* (2002), 325–337.

[2] Ned Calonge, Diana B Petitti, Thomas G DeWitt, Allen J Dietrich, Kimberly D Gregory, Russell Harris, George Isham, Michael L LeFevre, Roseanne M Leipzig, and Carol Loveland-Cherry. 2008. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 149, 9 (2008), 627–637.

[3] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58, 3 (2011), 11.

[4] Emmanuel J Candès and Benjamin Recht. 2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics* 9, 6 (2009), 717.

[5] Sarah Stark Casagrande, Catherine C Cowie, and Judith E Fradkin. 2013. Utility of the US Preventive Services Task Force criteria for diabetes screening. *American Journal of Preventive Medicine* 45, 2 (2013), 167–174.

[6] Laura Davisson. 2016. USPSTF breast cancer screening guidelines. *West Virginia Medical Journal* 112, 6 (2016), 29–32.