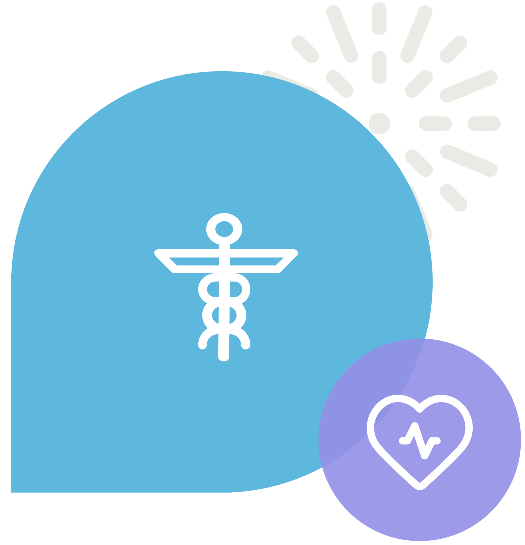
Predictive models for healthcare organizations

Promote strategies for acquisition, timely intervention, and population health.

WebMD Ignite™



As the healthcare landscape continues to change, healthcare organizations can harness consumer data and modern analytics to develop sound strategies. Use predictive models to maximize your market share, manage costs, strengthen physician relationships, and improve patient care.

How can healthcare organizations benefit from these fundamental changes? Healthcare predictive models are one solution. When designed and implemented using best practices, these models can empower healthcare organizations to intelligently acquire new patients, manage populations by providing the right care at the right time, and increase alignment and loyalty from affiliated providers.

Healthcare has undergone a radical transformation with reimbursements, hospital size, and provider employment. And now consumers, who had few healthcare choices in the past, are making informed decisions and driving their own care. Predictive models allow healthcare organizations to plan for the future to maximize market share, improve patient care, and manage costs.

1

Understanding predictive models

What value do healthcare-specific predictive analytical models provide? They tell healthcare organizations exactly who their patients are and who may need their services soon. Standard analytics can tell them what happened in the past. Monitoring can tell them what is happening now. But predictive analytics can provide a glimpse into what may happen in the future — so healthcare organizations can plan for it.

Predictive models are:

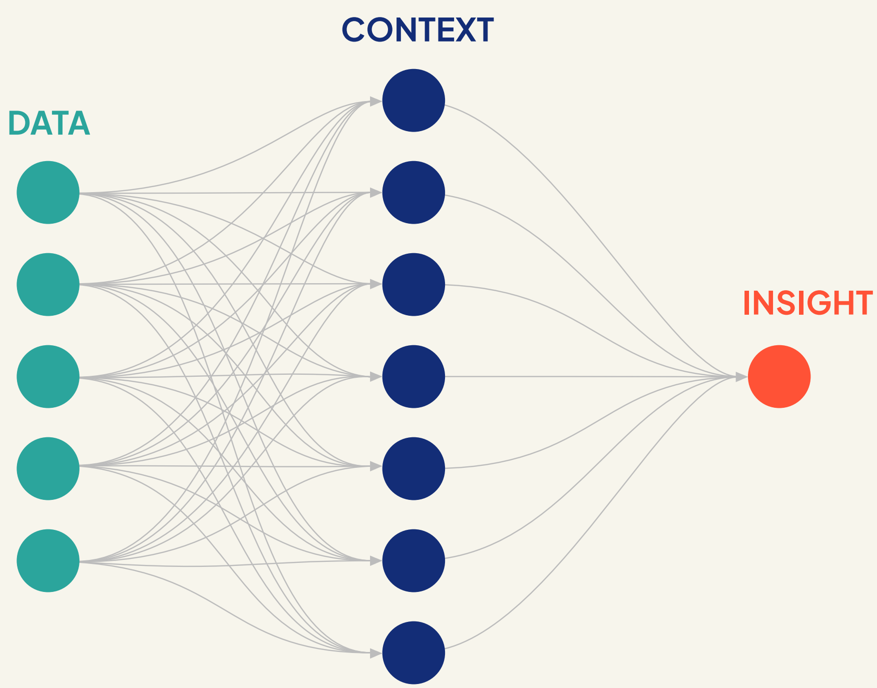
- Statistical formulas designed to capture trends and relationships between variables. Predictive models are based on detailed, collected data. They are validated and revised as additional data becomes available.
- Mathematical methods can best analyze a particular problem. Sophisticated development methods produce a simple outcome that is easy to understand.
- Tools used to predict future behavior. In healthcare, predictive models forecast the likelihood of a future health event for individuals. Hospitals can then project future population health needs for strategic planning and communication.

Examples of predictive model applications:

Predictive models solve problems in a wide variety of contexts. Credit card companies can flag suspicious buying behavior to detect fraud. Consumer websites like Netflix and Amazon can make highly relevant, personalized recommendations. Predictive models revolutionized direct marketing by allowing managers to focus on the individuals most likely to make a purchase.

In healthcare, predictive modeling adds value by identifying individuals most likely to have medical needs. Neural networks are the most useful predictive modeling method for healthcare. Neural networks identify risk by capturing the uniquely complex interaction between healthcare utilization, demographics, medical codes, and visit history.

PREDICTIVE MODELS SOLVE PROBLEMS IN A VARIETY OF CONTEXTS



How do neural networks work?

A neural network is a mathematical model that converts input values to an output score through a process called artificial learning. Four key attributes make neural networks effective at understanding health utilization:

1. **Scalability:** Neural networks quickly score large data sets, allowing a score refresh with each database update.

2. **Focus:** Neural networks automatically identify important variables and ignore noise. For example, if RV ownership does not impact diabetes risk, the neural network ignores an "RV owner" input variable to avoid over-fitting.

3. **Identification of nonlinearities:** Many healthcare relationships are complex. For example, aging from 20 to 30 years old has only a small impact on the risk of heart disease relative to aging from 70 to 80 years old.

4. **Recognition of interactions:** The effect of some variables can be enhanced or mitigated by other variables. For example, heart disease risk increases with age more quickly for men than for women.

2

Building a smart model

All patients are unique, and smart predictive models can help identify specific health risks and needs. A healthcare organizations database might contain a patient — we'll call him John Smith — who is a sedentary, 55-year-old man. John might be considered a typical cardiac patient, given his age and aversion to exercise. His younger wife, Susan (also in the database), is 40 and an avid runner. She doesn't appear to be an obvious candidate for cardiology services.

Surprisingly, Susan is more likely to need cardiology services than her husband. This insight is possible because her predicted risk scores for heart-related diseases are much higher than John's. The data reveals that Susan has a family history of heart disease and was recently diagnosed with hypertension.

Patients are not typical and do not always fit a stereotype or persona. Smart predictive modeling enables healthcare organizations to find individuals who need health services. It finds at-risk audiences, so physicians can provide appropriate patient diagnosis and care.

Legacy model: cluster codes

In the past, predictive modeling in healthcare was difficult due to a lack of comprehensive and historical patient data. Prior to using predictive models, marketers, statisticians, and clinicians used cluster codes to find prospects.

Cluster codes place households into cohorts sharing a set of socioeconomic characteristics. A cluster will typically span a range of ages and incomes, and possibly ethnicity, urbanization, and patterns of consumption. Cluster assignment is based on geo-coded address, with this organizing principle: Someone who lives in the same neighborhood as households with known characteristics probably shares those characteristics. Cluster codes use very small geographic divisions such as census blocks and six-digit postal code extensions. Clusters are given creative names such as "Milk and Cookies" or "Shotguns and Pickups" to evoke images of their lifestyles and associated economic behaviors.

Cluster codes perform well for general, non-medical household data based on zip codes, buying patterns, and generic information. But they don't enable healthcare organizations to understand a population's health situation or identify individuals who may soon need certain services.

Consumer model

A consumer model is designed to run in an environment in which access to health records is not available. It uses market demographics to predict future health needs for both patients and non-patients.

Every patient in a healthcare organizations' database should receive a set of scores from multiple (100 or more) patient models. In addition, patients and prospective patients should also receive a set of scores from multiple consumer models, which can be recalculated monthly for each individual. Including a geocoded component provides U.S. Census information about the neighborhood where the individual lives.

Healthcare marketers and planners can create consumer and patient models for nearly all service lines and some sub-service line specialties. Some models can also target utilization by encounter type: inpatient, outpatient, or emergency services.

Consumer models are trained using only the demographic data for patients who have used medical services for the procedures or diagnoses targeted by the model. When a consumer model assigns a risk score to individuals, it is based on how closely their demographics resemble the patient the model was trained to recognize.

The solution for each of these individuals is more targeted and economical than cluster coding or random selection. Consumer models serve as predictors for the kinds and quantities of disorders and diseases in a market, helping healthcare organizations to optimize strategic plans by identifying who needs what health service.

Healthcare predictive models	Cluster codes
Based on healthcare variables and predictive algorithm	Based on non-healthcare variables and clustering algorithms
Segment market based on differences	Segment market based on similarities
Predict individual service use	Predict group/family behavior
Can be based on millions of encounters	Based on 10 variables
Scores based on ICD-10, MS-DRG, CPT categories	Scores based on single market model
Provide multiple individual scores when consumers and patient models are combined	One score per family
Dynamic, updated, and integrated with EHR data	Static, stale data that is not integrated with EHR data

Patient model

With the widespread adoption of electronic health records (EHRs), healthcare organizations can now obtain an accurate picture of their patient population and make decisions based on real data, not assumptions. A patient model assigns medical utilization risk, taking into account an individual's medical history. It examines recency, frequency, type, and service line of a patient's medical visits.

Ideally, the model should use hundreds of demographic data points and all available medical records to predict patients' future health needs. It should consider the same demographic variables used in the consumer model, as well as codes for chronic conditions, personal and family history, evaluation and management, and medical imaging.

The patient model should be trained using the coded medical histories of patients who have used medical services for targeted procedures or diagnoses. It can then assign risk scores to ensuing patients based on how closely their medical history resembles the patients it has been trained to recognize.

In the examples on the following page, healthcare organizations can predict what health condition(s) a current or past patient has, and assign them a risk score to provide a sense of urgency. The scores are based on the output of what the models identify about each person in the database.

A patient model scores patients higher or lower depending on events in their medical history. Medical codes that appear more frequently prior to the target utilization raise the score, while medical codes appearing less frequently lower the score. The model should also be sensitive to code combinations seen more frequently prior to a visit.

The examples below illustrate the types of data that feed into the models. The scores are based on output of what the models identify about each person in the database.

EXAMPLE 1

DEMOGRAPHICS
- Male
- Age 45
- Married
- Children present
- Median household income
- Zip code

CONSUMER SCORES
- Medical cardiology inpatient score: 731
- Diabetes outpatient score: 773
- Knee replacement: 568

EXAMPLE 2

DEMOGRAPHICS
- Female
- Age 60
- Divorced
- No children present
- High household income
- Zip code

CONSUMER SCORES
- Medical cardiology inpatient score: 537
- Diabetes outpatient score: 443
- Knee replacement: 555

EXAMPLE 3

DEMOGRAPHICS
- Female
- Age 30
- Single
- Children present
- Low household income
- Zip code

CONSUMER SCORES
- Medical cardiology inpatient score: 324
- Diabetes outpatient score: 394
- Knee replacement: 179

Patient models can predict what health condition(s) a current or past patient has or may develop, and assign that patient a risk score to provide a sense of urgency.

Patient has normal cholesterol, diagnosed with patellofemoral pain syndrome (runner's knee)

EXAMPLE 1

DEMOGRAPHICS
- Male
- Age 45
- Married
- Children present
- Median household income
- Zip code

PATIENT SCORES
- Medical cardiology inpatient score: 628
- Diabetes outpatient score: 674
- Knee replacement: 832

Patient has personal history of tobacco use, diagnosed with benign essential hypertension (elevated blood pressure)

EXAMPLE 2

DEMOGRAPHICS
- Female
- Age 60
- Divorced
- No children present
- High household income
- Zip code

PATIENT SCORES
- Medical cardiology inpatient score: 813
- Diabetes outpatient score: 782
- Knee replacement: 686

Patient has family history of heart disease, diagnosed with impaired fasting glycemia (prediabetes)

EXAMPLE 3

DEMOGRAPHICS
- Female
- Age 30
- Single
- Children present
- Low household income
- Zip code

PATIENT SCORES
- Medical cardiology inpatient score: 660
- Diabetes outpatient score: 627
- Knee replacement: 327

Interpreting consumer and patient model scores

Consumer models and patient models should be created using several years of pooled de-identified data, preferably from multiple healthcare organizations that present a cross-section of the national population. An individual's medical history then allows consumer and patient model scores to be generated at a specified point in time. Healthcare organizations can measure model performance by examining the individual's medical history following the score.

Score

Each individual should receive a score of 0 to 999 for each consumer model and patient model. These scores represent risk in an actuarial sense, meaning a relative abstract likelihood of the targeted event occurring within the next 12 months. A score of 800 indicates greater risk than 400, but not necessarily twice the risk; it simply serves as a metric that sorts individuals according to risk. It is important to note that relative risk is not the same thing as probability. Actual probability of the event occurring for a particular score depends on the utilization rate at which the event occurs in the population, and on the predictive power of the model. The term "risk" can also apply to positive events (for example, obstetrics patients and foundation donors).

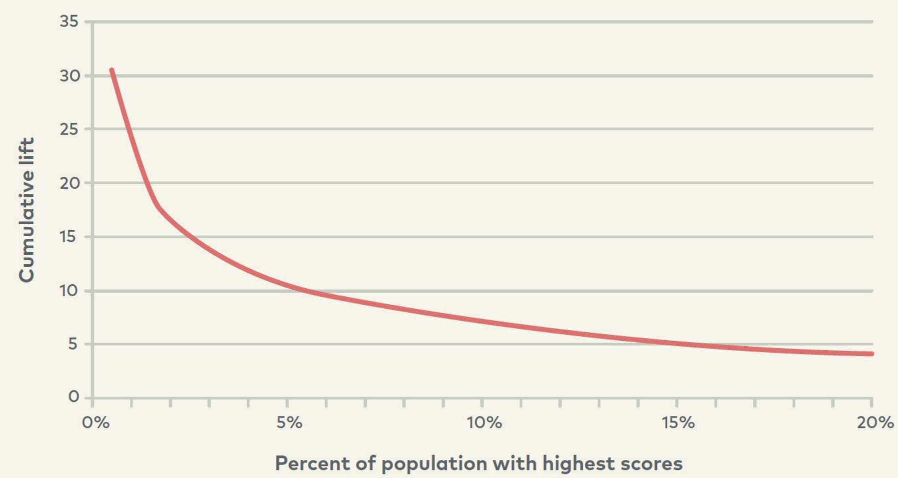
Lift

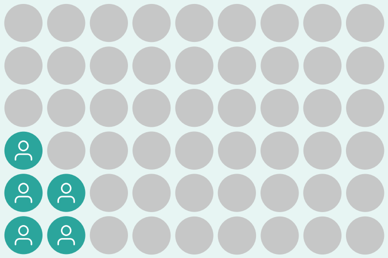
Lift (and cumulative lift) is a useful metric to convey the predictive power of a model [Figure 1]. It is the factor multiplied by the population utilization rate to produce the rate of utilization for a given score.

Cumulative lift is assessed over a sorted or classified interval of population. A model with a cumulative lift of 5 at 900 means that individuals with a score of 900 or higher have the targeted medical event 5 times as often as the general population. Describing predictive power in terms of lift as a multiplying factor removes variation in population utilization rate for differing medical events.

Lift extends directly to campaign planning. If 10% of the population falls in the 900+ score range, then for the cost of messaging 10% of the population, a campaign will each 5 x 10% = 50% of the individuals who may have the targeted medical event in the next year. This estimate is possible without knowing the exact service utilization rate. A known utilization rate permits estimates of specific numbers of individuals and medical encounters.

FIGURE 1: CUMULATIVE LIFT MEDICAL CARDIOLOGY

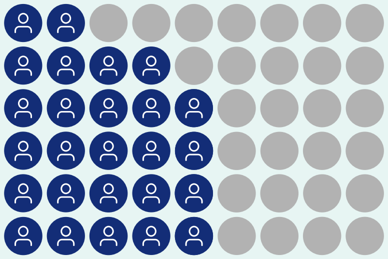




Picking people **randomly** (no models) has a lift of **1x**.

VS.

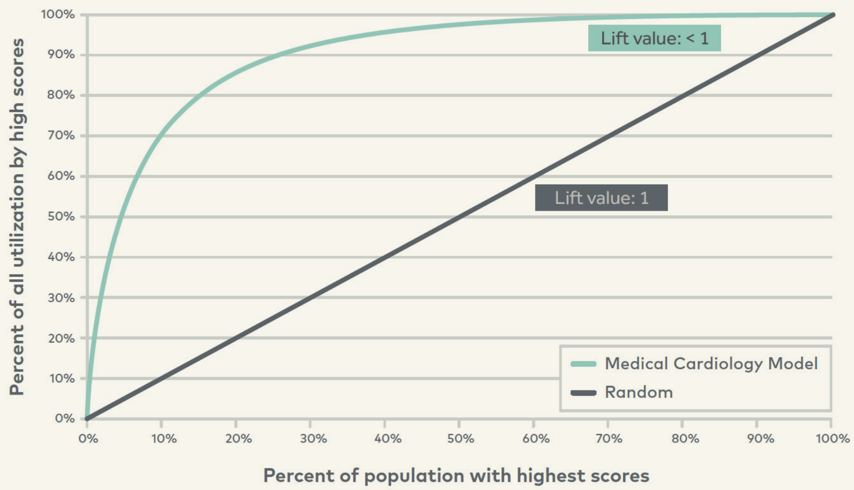
Models with a lift of **5x** are **five times better** at getting to the right person than picking at random.



Utilization curve

A utilization curve (also known as a cumulative response curve) extends the concept of lift [Figure 2]. It shows the cumulative percent of the total population using services for the targeted medical event (vertical) for the cumulative percent of population as ranked by score (horizontal). Lift is the slope of the curve on the graph. A diagonal line has a lift of one, and is equivalent to using no model, or to randomly selecting individuals using no criteria. On the left there is high lift, where the curve rises sharply. Where the curve rounds out, there is low lift. Where the curve is flat near the top of the chart, there is a lift value lower than one. Messaging individuals in the flat range of the utilization curve is counterproductive because utilization is concentrated in the high-scoring population. Reading a utilization curve gives a quick sense of the ROI that a model provides.

FIGURE 2: PATIENT UTILIZATION MEDICAL CARDIOLOGY



3

Using predictive models strategically

The future of healthcare is hardly certain. CMS programs for value-based purchasing and meaningful use of information technology, combined with other Affordable Care Act provisions and a shifting quality focus, make for a confusing path forward.

Predictive models can help healthcare organizations set and meet goals in the face of uncertainty. They indicate priorities in consumer, patient, and physician engagement to increase satisfaction and improve health outcomes. Knowing individual medical risks empowers economical, data-driven strategies to address them through preventive care and timely interventions.

Proactively manage population health

Predictive models identify and stratify individuals within populations, enabling actions that can save lives. Whether someone is at high risk or is already diagnosed with one or more diseases, healthcare organizations can target an intervention and guide that person to the most suitable healthcare provider. They can also plan for utilization and staffing in risk areas by identifying physicians for increased alignment or loyalty. By focusing on those who need it most, healthcare organizations can deliver higher-quality care and ultimately improve outcomes.

Engage at-risk patients to reduce costs

While healthcare organizations can easily identify high-risk patients from their medical records, predictive models empower them to find moderate-risk individuals — both patients and prospects — before they become high-risk. These people benefit most from proactive communications that provide education and encourage a doctor's visit. Strategic use of consumer and patient models for this purpose combines the predictive model with filter criteria to select patients and prospects with high scores who have not already had a major procedure or diagnosis. When healthcare organizations identify and engage these individuals, they can lower their direct costs of care by preventing major medical events.

Growing market share still matters

Growing market share is important to the financial health of many healthcare organizations. Consumer models can map utilization versus risk throughout a service area. Healthcare organizations can then message prospects in areas where consumer model scores indicate higher potential demand for healthcare services. Used strategically, predictive models can help providers offer preventive care, proactively engage patients, and take positive steps toward population health, increased care quality, and proven ROI.

④

Optimizing model strength

- **Comprehensive:** All available data should be collected to create the most accurate and complete view of the market and patient population.
- **Clean:** Legacy healthcare IT has created a "best-of-breed" approach, meaning there are few places in which integrated data exists. Often data accuracy is less than optimal.
- **Actionable:** Data without insight is just noise. Disparate information sources (information silos) must be compiled into a centralized database, and scrubbed prior to analysis, to avoid duplication and other inconsistencies. This is a highly organized process that transforms information once designed for specific purposes into a flexible scientific data powerhouse. For examples, to gain actionable insights, healthcare organizations can use a full demographic data set lined up next to a list of medical histories.
- **Accurate:** To ensure the database has the best possible information for modeling and reporting, data from a wide variety of sources should be incorporated and linked.
- **Robust:** The database must be tailored to represent an individual healthcare market based on specific data sources. Healthcare-specific models should contain hundreds of data sources and represent as much of a given population as possible. Market data points need to be matched to medical records.

Predictive models should be benchmarked against other healthcare organizations' medical data when possible. Healthcare organizations can continually improve the quality and accuracy of their models by comparing predicted outcomes to actual outcomes.

Better-informed models mean better predictions and better health and business outcomes.

5

Case study: A test of WebMD Ignite predictive models

How effective are WebMD Ignite consumer and patient models? We performed a comparison test against two other methods commonly used: cluster codes and expert queries.

Consumer model vs. cluster codes

Though cluster codes are designed for general purposes of marketing and not specifically for healthcare utilization, certain clusters may use certain medical services at higher rates than other clusters. To apply cluster codes to healthcare customers, WebMD Ignite first cluster-coded patients and prospects in a given service area. Then, we performed a detailed analysis of utilization by service line to establish the variation in utilization rates. At this point, lift could be determined for each service and cluster combination. The clusters were ranked from high to low and selected for messaging according to their lift or utilization rate.

While some clusters showed increased lift for certain medical events, clusters typically comprised a small percentage of the population because the population is typically divided into 60 or more clusters. Several clusters needed to be combined to create a campaign that reached a significant fraction of the utilization of the targeted service.

WebMD Ignite performed a test of cluster codes versus consumer and patient models for the purpose of messaging patients and prospects for selected service lines (medical cardiology, diabetes, joint replacement, obstetrics, and emergency room non-compliance). Scores for the consumer model and patient model were generated for 1.4 million patients and 4 million prospects in the combined service area of a large multi-hospital network using the date January 1, 2021. Patient medical history prior to 2021 and utilization in 2021 was available for the test, as was demographic information for all patients and prospects in the service area. Addresses were geo-coded and clusters were applied. Because cluster codes apply to a household, when a model selected a household for messaging, credit was given if anyone in the household utilized the targeted service.

Based on the actual behavior of the people in the study, lift and utilization curves were generated to show the comparative performance of the models. In all comparisons, the consumer and patient models exceeded the performance of the cluster codes. Since cluster codes do not have access to a patient's medical history when a cluster is assigned, the discussion of cluster-code performance in this section pertains only to consumer models.

For medical cardiology, the top eight performing clusters comprising about 2% of the population of the service area were "Rural Bypasses," "Social Security Net," "Modest Income Homes," "Urban Rows," "City Dimensions," "City Commons," "Simple Living," and "Pacific Heights." The lift of these combined basic cluster codes was 4.1. Their combined utilization accounts for about 8% of total medical cardiology services. By comparison, using the more specific consumer models, the top 2% of the population has a medical cardiology lift of 5.5, accounting for 12% of the total medical cardiology utilization [Figure 3].

Going deeper into the population, the cluster approach takes the top 22 clusters combined to find 38% of the utilization in 17% of the population with a cumulative lift of 2.3. At the same list size, the consumer model achieves a lift of 3.6 and accounts for 61% of the medical cardiology utilization [Figure 4]. All services examined experienced similar performances. The cluster codes begin to lag behind the consumer model in the most at-risk segments. The performance gap steadily widens as the reach exceeds 20% of the population.

FIGURE 3: MEDICAL CARDIOLOGY UTILIZATION CURVE

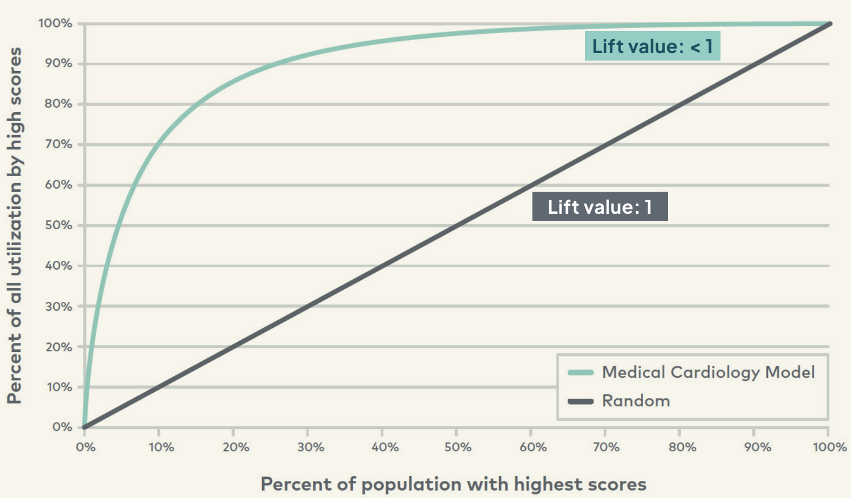
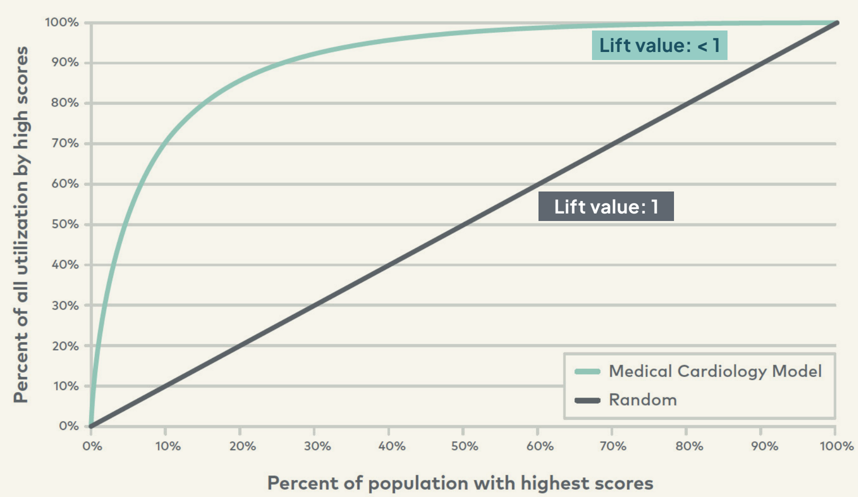


FIGURE 4: MEDICAL CARDIOLOGY UTILIZATION CURVE (ZOOMED 20% REACH)



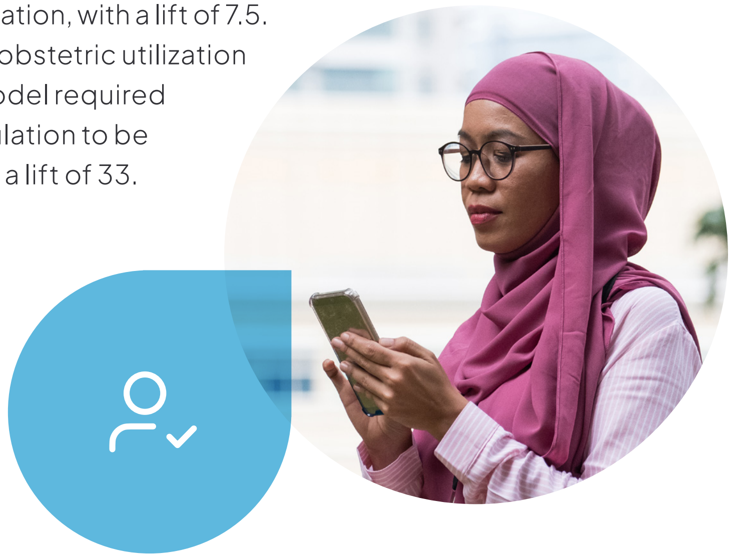
The best-performing cluster and service-line combination was a lift of 10.6 for "Dorms to Diplomas" for emergency room non-compliance. Unfortunately, only 0.6% of the population in the service area was coded with this cluster. A cluster named "Prosperous Empty Nesters" also had a lift of 0.9 for obstetrics. This signifies that they utilized obstetrics services at a slightly lower rate than average for the service area. The lift indicated that they should not be targeted in an obstetrics campaign, but we know there are women giving birth within this higher-income cluster, and these prospects will be overlooked in a cluster-based obstetrics campaign. A consumer model indicated the specific households where age and presence of children combine with other variables to suggest a higher likelihood of utilization.

There are two principal flaws in using cluster codes for healthcare marketing. The first flaw is that the clusters are not designed based on the characteristics of patients utilizing medical services. Consumer and patient models are essentially clusters based specifically on utilization. It is far more effective to choose a consumer model-based group of households defined by the service line "Medical Cardiology" than it is to piece together several clusters defined by the patterns of consumption of retirees.

The second flaw is that membership in a cluster is exclusive. If your household is labeled "Milk and Cookies," you cannot also be labeled "Heartland Communities" or "High-Rise Renters." Therefore, some clusters will be selected at the cost of excluding others. In addition, many excluded clusters contain individuals with a targeted risk, but they are the minority population within a cluster dominated by people with lower risk. Assigning someone a high risk for diabetes prevents assigning that same person a high risk for heart disease. With consumer and patient models, every person receives scores for every model. The highest-risk condition does not conceal other conditions with elevated risk. It is often very useful to investigate populations at risk for multiple related conditions.

Consumer and patient models vs. expert queries

An expert query is another tool available to healthcare marketers who have access to patient medical records. For certain service lines, good lift can be obtained by only using a few criteria. As a basic example of an expert query, WebMD Ignite created a list for obstetrics utilization based on a target female between ages 24 to 39. This set of individuals comprised approximately 12% of the service-area population, and accounted for 81% of all obstetrics services utilized, for a lift of 6.3. At 12% of the population messaged, an expert query found 27% of the utilization, with a lift of 2.2. The consumer model found 98% of the utilization, with a lift of 7.5. Finding 81 percent of obstetric utilization with the consumer model required only 2.3% of the population to be messaged, achieving a lift of 33.



Expert queries can be combined with other models, such as clusters and consumer models. An expert query can be used as an exclusion criterion to filter selected individuals and households from the campaign. An expert query can also be used as an inclusion criterion to ensure that selected individuals and households are included in the campaign, regardless of inclusion by the model.

In our obstetrics example, when we used the expert query (female aged 24 to 39) as an inclusion criterion, it filtered the prospects to 12% of the service area in which 81% of the utilization occurs. When cluster codes were applied to the filtered population, there was a small gain due to the few clusters performing better than 6.3 lift, but this occurred only for a fraction of a percent of the population. The filtered cluster found 70% of the obstetrics utilization in 8.6% of the population, an 8.1 lift. The consumer model did not get a significant boost from the expert criteria. The inclusion criteria constrained the consumer model to selecting no more than 81% of the utilization because, while small, the remaining obstetric utilization occurred for women outside the 24 to 39 age range. The consumer model found most of this utilization by considering other demographic variables.

Using the expert query as an inclusion list added no performance to the obstetrics consumer model because the consumer model was already trained to select nearly every one of the prospects that are selected by the filter. The combination of the inclusion list and cluster codes performed no better than the filter alone because the filter selected nearly every one of the prospects coded by the highutilization clusters.

Our obstetrics example was limited to prospects, and it used criteria that did not require medical expertise to design. When campaigns are limited to the patient population, marketers can select the campaign target based on diagnosis and prior visits. These criteria will require medical expertise and knowledge of medical coding. As an exclusion filter becomes more detailed, it restricts the list to a smaller number of patients. An expert query can be elaborately designed to conditionally include patients based on a tree of decisions. Designing a set of advanced expert criteria is limited only by the data available and the resources with which to analyze it.

WebMD Ignite has found that advanced expert queries can perform at a level comparable to a patient model for specific conditions and procedures within a service line. It is, however, very difficult for a human to write a comprehensive set of criteria to achieve patient model lift for a more generally defined service line combining a variety of related procedures and conditions. Our experience working with consumer and patient models has shown that a simple set of expert criteria can augment the performance of a predictive model when applied either as an exclusion filter or inclusion list pertaining to specific diagnoses and procedures.

Using an expert query to match the predictive power of WebMD Ignite consumer and patient models requires considerable resources, and means essentially developing hospital-specific models. However, expert queries do consistently outperform cluster codes, especially when patient data is available.

6

Conclusion

Armed with powerful consumer and patient data models, healthcare organizations can reach the right audience with the right message via the right channel at the right time.

Predictive models are advanced mathematical techniques that can be used to more accurately identify individuals in the marketplace based on their health status. They are superior to cluster codes in selecting appropriate patients and consumers for education, disease management, and intervention programs. Predictive modeling is more efficient in reaching the right individuals with the right message, so hospitals can ensure they deliver the right care.

Using predictive-model best practices improves targeted population health by helping consumers and patients make healthier choices.

About WebMD Ignite

WebMD ignite is a technology and data analytics company that empowers healthcare organizations to engage consumers and optimize provider relationships to accelerate growth.
Our customers benefit from 30 years' experience applying data analytics to drive intelligent engagement and enable personalized healthcare journeys. At WebMD Ignite, we help healthcare organizations create seamless consumer experiences and improve outcomes to build healthier communities.

WebMD Ignite™

Unlock market dynamics and top growth opportunities.

Get started at webmdignite.com.