JPMSA
JOURNAL OF THE PHARMACEUTICAL MANAGEMENT SCIENCE ASSOCIATION

SPRING 2019 | IN THIS ISSUE:

pmsa
PHARMACEUTICAL MANAGEMENT SCIENCE ASSOCIATION

Table of Contents

Official Publication of the Pharmaceutical Management Science Association (PMSA)

The mission of the Pharmaceutical Management Science Association not-for-profit organization is to efficiently meet society's pharmaceutical needs through the use of management science.

The key points in achieving this mission are:

- Raise awareness and promote use of Management Science in the pharmaceutical industry
- Foster sharing of ideas, challenges, and learning to increase overall level of knowledge and skill in this area
- Provide training opportunities to ensure continual growth within Pharmaceutical Management Science
- Encourage interaction and networking among peers in this area

Please submit correspondence to:
Pharmaceutical Management Science Association
446 E High Street, Suite 10
Lexington, KY 40507
info@pmsa.net
(877) 279-3422

Executive Director:
Stephanie Czuhajewski, CAE
sczuhajewski@pmsa.net

PMSA Board of Directors

The *Journal of the Pharmaceutical Management Science Association (ISSN 2473-9685 (print) / ISSN 2473-9693(online))* is published annually by the Pharmaceutical Management Science Association.

PMSA Journal: Spotlighting Analytics Research

Welcome to the seventh edition of the *Journal of the Pharmaceutical Management Science Association (PMSA)*, the official research publication of PMSA.

The purpose of the Journal is to promote and embody the mission of the association, by:

- Raising awareness and promoting the use of Management Science in the pharmaceutical industry
- Fostering the sharing of ideas, challenges, and learning to increase the overall level of knowledge and skill in this area.

The Journal publishes manuscripts that advance knowledge across a wide range of practical issues in the application of analytic techniques to solve Pharmaceutical Management Science problems, and that support the professional growth of PMSA members. Articles cover a wide range of peer-reviewed practice papers, research articles and professional briefings written by industry experts and academics. Articles focus on issues of key importance to pharmaceutical management science practitioners.

If you are interested in submitting content for future issues of the Journal, please send your submissions to info@pmsa.net.

GUIDELINES FOR AUTHORS

Summary of manuscript structure: An abstract should be included, comprising approximately 150 words. Six key words are also required. All articles and papers should be accompanied by a short description of the author(s) (approx. 100 words).

Industry submissions: For practitioners working in the pharmaceutical industry, and the consultants and other supporting professionals working with them, the Journal offers the opportunity to publish leading-edge thinking to a targeted and relevant audience.

Industry submissions should represent the work of the practical application of management science methods or techniques to solving a specific pharmaceutical marketing analytic problem. Preference will be given to papers presenting original data (qualitative or quantitative), case studies and examples. Submissions that are overtly promotional are discouraged and will not be accepted.

Industry submissions should aim for a length of 3000-5000 words and should be written in a 3rd person, objective style. They should be referenced to reflect the prior work on which the paper is based. References should be presented in Vancouver format.

Academic submissions: For academics studying the domains of management science in the pharmaceutical industry, the Journal offers an opportunity for early publication of research that is unlikely to conflict with later publication in higher-rated academic journals.

Academic submissions should represent original empirical research or critical reviews of prior work that are relevant to the pharmaceutical management science industry. Academic papers are expected to balance theoretical foundations and rigor with relevance to a non-academic readership. Submissions that are not original or that are not relevant to the industry are discouraged and will not be accepted.

Academic submissions should aim for a length of 3000-5000 words and should be written in a third person, objective style. They should be referenced to reflect the prior work on which the paper is based. References should be presented in Vancouver format.

Expert Opinion Submissions: For experts working in the Pharmaceutical Management Science area, the Journal offers the opportunity to publish expert opinions to a relevant audience.

Expert opinion submissions should represent original thinking in the areas of marketing and strategic management as it relates to the pharmaceutical industry. Expert opinions could constitute a review of different methods or data sources, or a discussion of relevant advances in the industry.

Expert opinion submissions should aim for a length of 2000-3000 words and should be written in a third person, objective style. While references are not essential for expert opinion submissions, they are encouraged and should be presented in Vancouver format.

Industry, academic and expert opinion authors are invited to contact the editor directly if they wish to clarify the relevance of their submission to the Journal or seek guidance regarding content before submission. In addition, academic or industry authors who wish to cooperate with other authors are welcome to contact the editor who may be able to facilitate useful introductions.

Thank you to the following reviewers for their assistance with this issue of the *PMSA Journal:*

Eze K. Abosi, IPM.ai

George A. Chressanthis, Ph.D., Axtria

Mitch DeKoven, MHSA, IQVIA

Ira Haimowitz, Ph.D.

Srihari Jaganathan, UCB

Jukti K. Kalita, Pfizer

Gaurav Kapoor, Indegene

Ewa J. Kleczyk, Ph.D., Symphony Health Solutions

Sudhakar Mandapati, Strategic Research Insights, Inc.

Ashish C. Patel, CareSet Systems

JP Tsang, Ph.D. & MBA (INSEAD), Bayser

Editor: Igor Rudychev, AstraZeneca

UPCOMING PMSA EVENTS

2020 PMSA Annual Conference
May 17-20, 2020
Omni Hotel
Louisville, Kentucky

2021 PMSA Annual Conference
May 23-26, 2021
Omni Hotel
Nashville, Tennessee

Simple Probability Models for Predicting Aggregate or Sparse Data: An Empirical Analysis of Projecting Patient Persistency

Srihari Jaganathan, Head, Advanced Analytics, UCB Inc. and Ka Lok Lee, Senior Associate Director, Advanced Analytics and Consulting Services, IQVIA

Abstract: There are many instances in pharmaceutical analytics where there is a lack of availability of data to perform analysis and to gain insights. As an example, long term data on patient persistency is critical for constructing patient flow models in pharmaceutical forecasting. Persistency rates are typically available only for short at the *aggregate* level (e.g., persistency rates by month and product). A much longer duration of persistency rates is required in the analysis of patient flow models, usually 5 or 10 year horizons. It becomes necessary to forecast to the desired horizons by projecting patient persistency. Typically this is achieved by using simple curve fitting techniques in spreadsheets on aggregate persistency data. This would sometimes provide sub-optimal and irrational projections. Lee et al. proposed a very effective and simple probability based approach called the Beta-Geometric model (BG) to project patient persistency rates[1]. Fader et al. later proposed an extension called the Beta-discrete-Weibull model (BdW)[2]. The main objective of this work is to empirically analyze these models on persistency data from diverse disease states such as RA, epilepsy, osteoporosis, immunology, statin, hypertension and others. We discuss the behavioral insights gained from these models with implications to understand the patient persistency data. We further demonstrate the ease and advantages of using our recommended models over the current widely used practice of curve fitting to project persistency rates. All computations are performed using Microsoft® Excel® and R statistical software package.

Keywords: Drug persistency, Forecasting, Beta-Geometric, Probability models, Beta-discrete-Weibull, Latent class

Introduction and Motivating Example

There are many instances in pharmaceutical analytics where there is *lack* of availability of data to perform analysis and to gain insights. As an example, long term data on patient persistency is critical for constructing patient flow models in pharmaceutical forecasting. Persistency rates are typically available only for a short duration such as 12 months or 24 months due to availability of data. Moreover, the data sometimes are available only at the *aggregate* level. Figure 1 presents the persistency rates of bisphosphonates for 24 months[3]. A much longer duration of persistency rates is required in the analysis of patient flow models, usually 5 or 10 year horizons. It becomes necessary to forecast to the desired horizons by projecting patient persistency. Typically this is achieved by using simple curve fitting techniques in spreadsheets on aggregate persistency data. We apply 5 standard methods: Linear, Exponential, Log, Polynomial and Power to the data in an effort to project out the persistency curve. We use 8 months for testing/estimation, and the remaining 16 months for validation. The result is shown in Figure 1.

Lee et al. discussed three basic characteristics of a persistency curve and we find them to be worthwhile to be repeated here[1]. They are:

1. It starts out at 100% and works its way down towards 0% as time increases.
2. It is non-increasing.
3. It tends to decrease at a decreasing rate over time.

Figure 1: Persistency Rates of Bisphosphonates: Actual vs. Curve Fitting

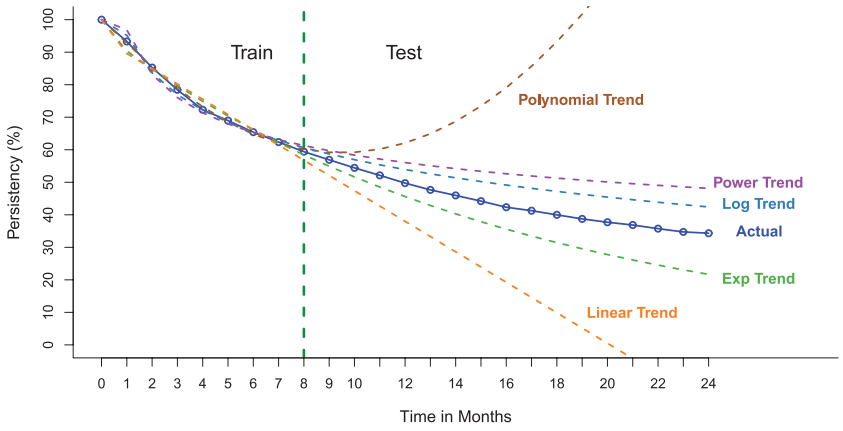
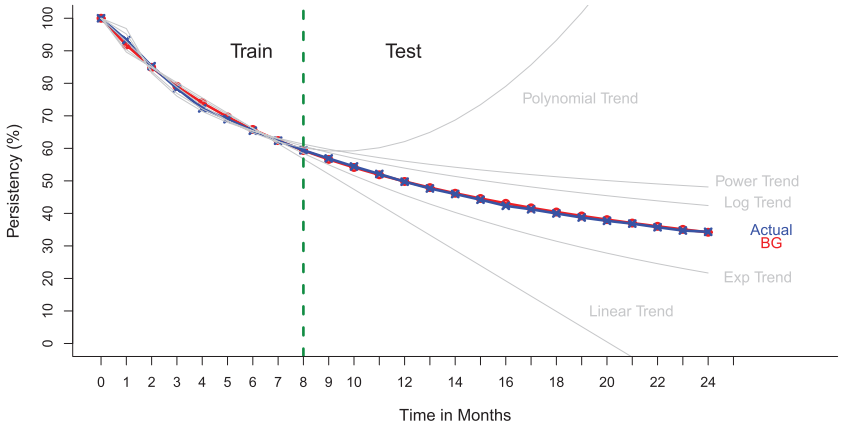


Figure 2: Persistency Rates of Bisphosphonates: Actual vs. Beta Geometric (BG)



Immediately we can see that only Exponential, Log, and Power methods can satisfy these basic requirements. Furthermore, while they may suffice from a simple curve-fitting purpose, they lack any behavioral story. Lee et al. proposed a very effective and simple probability model called the (shifted) Beta-Geometric model (BG) to project patient persistency rates[1], which we fit for same data in Figure 1. This is illustrated in Figure 2. The BG model fits the validation data like a glove and gives the best forecast. To the best of our knowledge, the adoption of the BG model for projecting persistency in pharmaceutical analytics is still limited, despite its simplicity and accuracy.

Fader et al. recently proposed extensions and variations to the BG model[2]. The main objective of this paper is to use an empirical meta-analysis to study these models on persistency data from diverse disease states such as diabetes, epilepsy, osteoporosis, immunology, statin and hypertension. The paper is organized as follows:

- Develop the methods here for the readers.
- Use empirics to show that these models forecast extremely well.
- Model selection strategies are provided.
- Conclude with future research.

Model Development

To project patient persistency, we apply three probability models that were developed for projecting customer retention rates in subscription businesses[2,4]. Even without individual-level longitudinal data, our modeling process considers the patients' refill decision at each cycle and accounts for suitable form of unobserved heterogeneity to capture the differences among patients. We show each model development in sequence and how they are related to each other. The data requirement needed to estimate these models are the same as the aggregate data used in our motivating example. For our forecasting approaches, we avoid trying to *explain* this important and repetitive decision to refill or stop medication, since individual factors are generally not captured in aggregate data and unknown in the forecasted periods. Finally, it is worth noting that the modeling techniques we are undertaking also belong to the class of the survival models[5]. Recognizing this relationship is helpful to understand these models.

Beta-Geometric Distribution (BG)

This model has already been applied to forecast patient persistency by Lee et al.[1]. We review it here because it remains a potent forecasting model and it lays the groundwork for the other two models we want to introduce.

As a mathematical construct, we consider whether a patient decides to refill a medication at the end of each cycle (usually each month) as a Bernoulli coin-flip decision with these assumptions:

1. If the coin comes up "heads", the patient refills the medication; if the coin comes up "tails", the patient does not refill and stops this medication altogether.
2. The coin does not change over time and each coin flip is independent of previous flips.
3. Everyone's coin is different.

The first two assumptions give us the well-known *Geometric distribution* to model a sequence of independent binary trials. Suppose the propensity to stop the refill is denoted by p. A patient refills t times and then stops the refill. The likelihood for this patient's data sequence is simply $p*(1-p)^t$. We can see that the conditional likelihood of refilling the medication is always $(1-p)$, regardless of the number of previous refills. Therefore, the Geometric distribution is a *memoryless* process. The corresponding survival function at period t, which is also the expression for persistency rate, is:

$$(1) \quad S(t \mid p) = (1-p)^t,\ 0 < p < 1,\ t \geq 0$$

The last assumption is to account for unobserved heterogeneity, since differences in patients' propensity to stop to refill exist. We choose to use the *Beta distribution* because it is flexible and conjugate to the Geometric distribution. The Beta distribution with parameters a and b, takes this form:

$$(2) \quad f(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)},\ a,b > 0$$

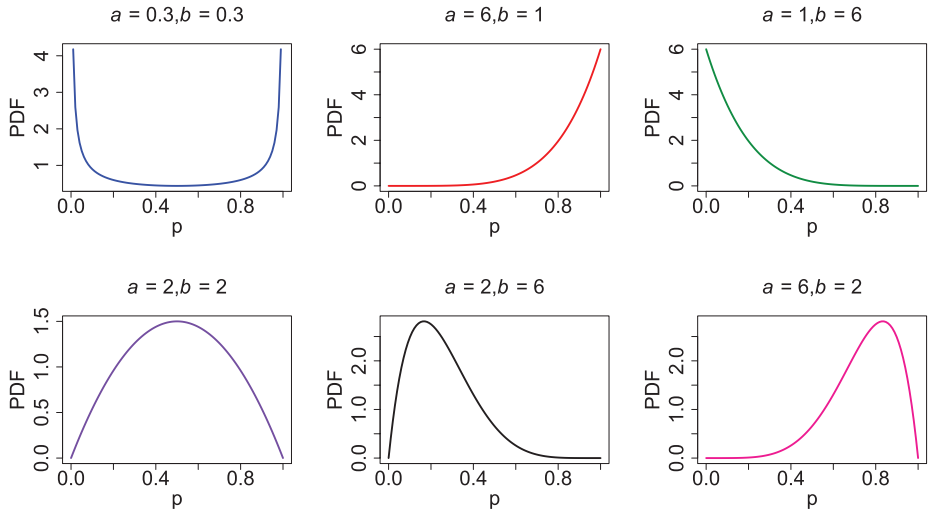
where $B(a,b)$ is the Beta function.

The Beta distribution can be used to understand the shape of refill propensities across the cohort patients by using the parameters values. Figure 3 shows the probability density function (PDF) of Beta distributions with different parameter values for a and b.

The mean of the Beta distribution is $\frac{a}{a+b}$. Another interesting measure is the polarization index, which is $\frac{1}{1+a+b}$. The polarization index is between 0 and 1, with 0 representing a homogenous cohort and 1 representing a polarized cohort. These can be useful metrics to understand the patient cohort at hand.

We combine the Geometric distribution and the Beta distribution to form the *Beta-Geometric*

Figure 3. Probability Density Function of Beta Distribution for Different *a* and *b*



distribution by solving the standard Beta integral. It gives us the following closed-form expression for the Beta-Geometric survival/ persistency function:

$$(3)\quad S(t|a,b) = \frac{B(a,b+t)}{B(a,b)}, t \geq 0, a, b > 0$$

Beta discrete-Weibull Distribution (BdW)
Fader et al. point out that many people struggle with the memoryless property of the BG distribution (which is also our assumption 2) on the surface[2]. In the subscription context, the customer is believed to become more "loyal" the longer they retain the subscription service. In our context, many would reasonably expect that there should be some type of positive "momentum" (or "inertia") over time with more familiarity with the medication. For example, a patient should be more likely to continue refilling if the side effect is minimal and the medication produces desired benefits. As a result, the propensity to refill should *increase* over time. This is an appealing hypothesis, but it should be tested empirically. In order to accommodate this expectation and test the hypothesis, we follow Fader et al. and modify the *Geometric distribution* to a *discrete Weibull distribution* (dW).

Suppose again the propensity to stop the refill is denoted by p. The corresponding survival function at period t for the discrete Weibull distribution, with a new parameter c, is:

$$(4)\quad S(t|p,c) = (1-p)^{t^c}, 0 < p < 1, c > 0, t \geq 0$$

It is easy to observe if $c>1$, $S(t|p,c)$ for the dW is smaller than the survival function for the Geometric distribution, all else being equal. Conversely, if $c<1$, $S(t|p,c)$ for dW is greater. If the conventional belief is correct that there is a momentum effect with the medication, we should expect the c parameter to be less than 1. Finally, if $c=1$, the dW becomes the Geometric distribution. These attractive properties make the dW the perfect extension to the Geometric distribution. Furthermore, we can continue to use the Beta distribution to account for unobserved heterogeneity in p, the propensity to stop the refill. Taken together, we form the *Beta-discrete-Weibull distribution* (BdW). The corresponding survival/persistency function at period t is

$$(5)\quad S(t|a,b,c) = \frac{B(a,b+t^c)}{B(a,b)}, t \geq 0, a, b, c > 0$$

We refer the readers to Fader et al. for more analytical analyses on this model[2].

k-latent-classes discrete-Weibull Distribution (LCdW)

The last model we want to introduce takes a different direction. So far, we have only used the Beta distribution to account for unobserved heterogeneity in the propensity to stop the refill, *p*. This form of heterogeneity is continuous and unimodal, as we displayed earlier in Figure 3. However, other forms of heterogeneity, say, a bimodal one, cannot be accounted for by Beta heterogeneity. Furthermore, in the BdW, there is no heterogeneity in the *c* parameter, which affects the momentum effect. It is equally valid to argue that differences exist in the *c* parameter across the patients. In order to account for both types of heterogeneities, we use a k-latent-classes discrete-Weibull distribution (LCdW)[2]. Specifically, for a 2 latent-classes model with class weights w and $(1-w)$, the corresponding survival/persistency function at period t is:

$$(6)\quad S(t|p_1,p_2,c_1,c_2,w) = wS(t|p_1,c_1) + (1-w)\,S(t|p_2,c_2)$$

where $S(t \mid p_i, c_i)$ is the survival expression for the discrete-Weibull distribution.

Immediately, one should be able to observe that this latent-class modeling approach is extremely flexible. One can easily extend to multiple latent classes beyond 2. Furthermore, one can replace the underlying dW with another distribution. There is no theoretical limit on either of these changes. However, the more latent classes one opts to use, the more parameters needed to be estimated. Alternative specifications on the underlying distribution may also lead to difficulty for interpretation. As such, we recommend the 2 latent classes here for practical reasons.

Estimation Method

The log-likelihood (LL) function for estimation from the entire patient cohort with T periods of data is:

$$(7)\quad LL(\Omega|data) = \sum_{t=1}^{T} n_t \ln[P(t|\Omega)] + (N - \sum_{t=1}^{T} n_t)\ln[S(T|\Omega)]$$

where

Ω is the set of parameters for the particular model,

$S(t|\Omega)$ is the survival function for the particular model, which can be equation (3), (5), or (6),

$P(t|\Omega)$ is the probability mass function at period t, which is equivalent to $S(t-1|\Omega)-S(t|\Omega)$,

N is the size of the entire cohort,

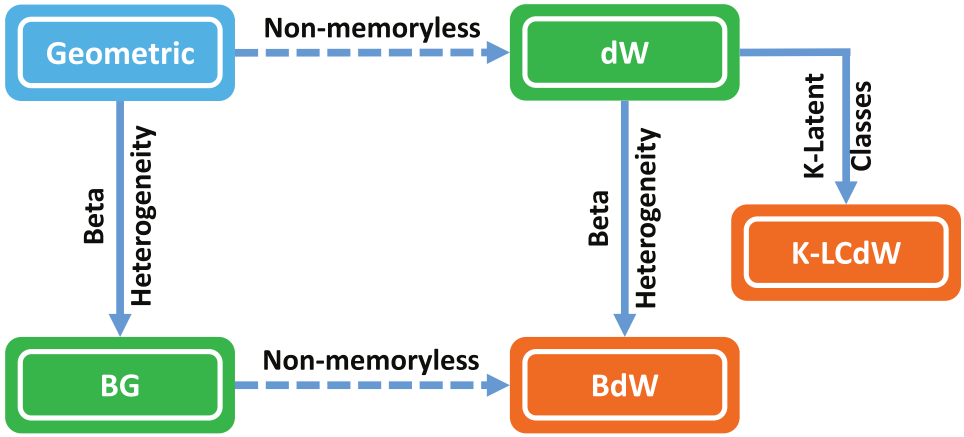
n_t is the number of people becoming non-persistent (i.e., stopping the medication) at period t.

We estimate the model parameters using the standard likelihood constructed from interval censored data for survival models[5]. In survival analysis terminology, the first term in the log-likelihood function is the contribution from the non-censored data. In our context, it is the data coming from the patients who have stopped the medication and become non-persistent in the observed period. The second term is the contribution from the censored data. These are the patients who are still refilling the medication at the end of the observed period.

We provide a graphical display of the relationships among different models we introduce in Figure 4. We use solid arrows to represent the heterogeneity assumptions and dash arrows to represent relaxation of the memoryless property. There are various pieces linking all the models together.

At this point, the readers may wonder why we chose these particular models, since there are countless other survival models available. We have provided the benefits of these models as we built them, but we summarize their benefits once again here. One main benefit is the computational convenience. The models are analytically tractable. For estimation purposes,

Figure 4. Relationships Between Different Models



we can easily program the log-likelihood function in a software with an optimization routine, even in a spreadsheet program such as Microsoft Excel with the Solver add-in to estimate the parameters. We can also estimate the parameters using other optimization routines available in R or other statistical software. Another important benefit is the behavioral interpretation that these models have. As we will show later, we can use the model parameters to give us diagnostics about the characteristics of the patients' behavior. Finally, we will also demonstrate the impressive predictive power of these models. All these benefits taken together are why we chose these particular models.

Empirical Meta-Analysis

In this section we describe the empirical analysis of BG, BdW and LCdW models compared with excel based trend models such as Linear (Lin), Exponential (Expo), Logarithmic (Log), Polynomial (Poly), and Power (Pow). Persistency data was extracted from published research in diverse therapy areas including Hypertension[6], Ocular Hypertension[7], Statin[8], Insulin[9], Epilepsy[10], Rheumatoid Arthritis (RA)[11], Osteoporosis[3], Alzheimer[12], ADHD[13] and Atrial Fibrillation[14]. The reason for choosing such a wide therapy area for empirical validation is to ensure that the models described in this article can be generalized. Persistency data from the aforementioned articles were extracted using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer) and discretized using R package akima[15].

We want to first show the results from another dataset, the Hypertension example in Figure 5. We can see that unlike the Bisphosphonates example, BG is not the winner. Instead, the winner is the LCdW. As such, we feel that it is imperative to perform the empirical meta-analysis to compare and contrast how these models behave more generally.

To assess the accuracy of the models, the extracted data was split into train and test datasets. Let the time period be denoted as $t = 1,2,\ldots x,x+1,\ldots T$. The train data, $t=1,2,\ldots x$, was used to build models and make forecast and test data, $t=x+1,x+2,\ldots T$, was used to assess the accuracy of the forecasted values compared with the actual values. Minimal observation values were used to train the models. For data that has longer time horizon we used the first 12 data points to train, while for shorter data, 5 or 6 observations were used to train the models.

Figure 5: Persistency Rates for Hypertension

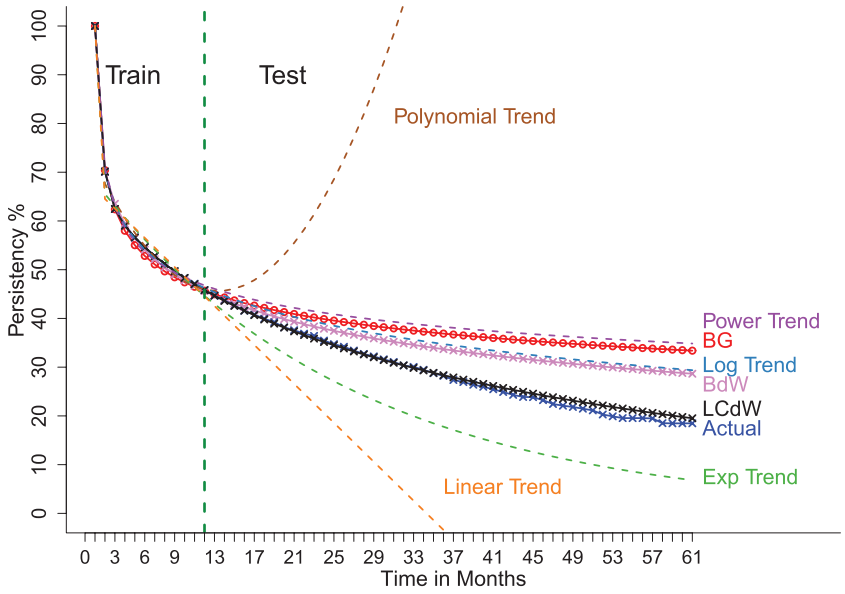


Table 1: Comparison of BG, BdW, LCdW vs. Excel Based Trend Models Using MAPE

Dataset	Train	Test	Lin	Expo	Log	Poly	Pow	BG	BdW	LCdW	Best
Hypertension	12	49	151.1	36.6	28.8	681.2	43.2	38.0	25.2	**3.3**	**LCdW**
Ocular Hypertension	5	7	102.7	31.2	9.5	197.9	16.2	5.1	10.6	**4.4**	**LCdW**
Statin	6	14	113.5	39.4	6.3	257.0	11.4	10.8	**1.0**	1.3	**BdW**
Insulin	12	19	26.9	17.3	2.9	16.8	5.6	**0.9**	4.6	2.9	**BG**
Epilepsy	6	13	97.6	18.1	25.0	10.2	43.8	**5.0**	11.5	9.6	**BG**
RA	12	49	102.3	32.7	23.2	136.9	31.6	23.7	**3.1**	24.2	**BdW**
Osteoporosis	6	19	80.8	24.9	18.5	226.0	28.2	**5.7**	17.3	37.8	**BG**
Alzheimer	6	7	26.4	13.4	4.6	19.8	9.6	**1.6**	2.2	3.5	**BG**
ADHD	12	61	61.2	11.7	50.3	25.0	54.6	**6.3**	21.4	46.6	**BG**
Atrial Fibrillation	6	13	48.9	21.4	7.3	101.8	13.9	**3.5**	9.7	23.0	**BG**

Accuracy measure Mean Absolute Percentage Error (MAPE) was used to compare the predicted values with actual values in the test data and it is given in equation 8.

$$(8)\quad MAPE = \frac{100}{(T-x)} \sum_{t=x+1}^{T} \frac{|Actual_t - Forecast_t|}{|Actual_t|}$$

All computations were performed using R statistical software package. The results using MAPE accuracy measure of Excel based vs. BG, BdW, and LCdW are reported in Table 1. The name of the dataset and the number of observations of train and test data are also shown in Table 1. Table 2 also provides the model parameters for all the datasets.

Empirical results indicate that in all 10 datasets, probability based models BG, BdW or LCdW significantly outperformed Excel based models. Figure 6 also illustrates the best model forecast fit for each individual dataset based on train and test. Considering that overall the BG, BdW and LCdW models perform very well, analysts should consider these models in their toolbox to forecast patient persistency.

Figure 6: Illustration of Forecast vs. Actuals of Best Models in Their Respective Datasets

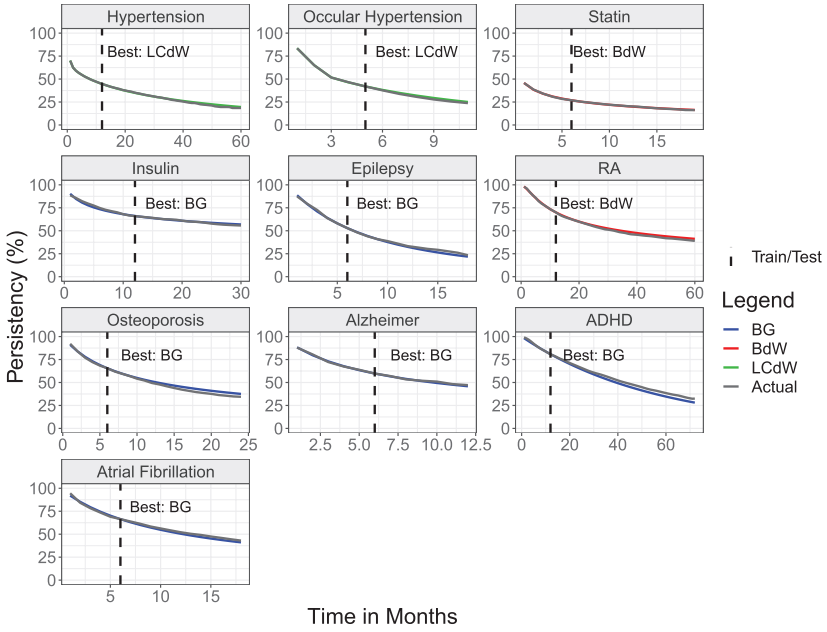


Table 2: Model Parameters for BG, BdW and LCdW

Dataset	BG		BdW			LCdW				
	a	*b*	*a*	*b*	*c*	*p1*	*c1*	*p2*	*c2*	*w*
Hypertension	0.2	0.4	1.1	2.6	0.4	0.819	1.023	0.096	0.625	0.280
Ocular Hypertension	1.5	6.2	0.2	1.0	2.4	0.030	4.947	0.193	0.717	0.172
Statin	0.3	0.2	6.4	5.4	0.3	1.000	0.626	0.542	0.290	0.001
Insulin	0.2	1.6	232.9	1890.2	0.5	0.086	1.166	0.127	0.001	0.315
Epilepsy	2.1	16.4	236.1	1790.7	0.9	0.009	3.315	0.139	0.714	0.102
RA	5.4	172.4	0.2	13.3	1.5	0.053	1.829	0.014	1.158	0.104
Osteoporosis	0.6	6.1	0.2	2.1	1.6	0.197	1.215	0.001	0.626	0.401
Alzheimer	0.6	4.2	0.7	5.2	0.9	0.151	1.317	0.107	0.434	0.305
ADHD	109.2	6097.3	0.2	24.6	1.6	0.027	1.539	0.001	1.013	0.255
Atrial Fibrillation	0.8	9.5	0.1	1.4	2.7	0.165	1.447	0.001	0.665	0.366

While it is not the focus of this paper to discuss each therapeutic category tested in depth, we want to discuss a few of them to illustrate how to gain insights about the patient cohort from the model parameters. Take the Statin category as the first example. The BdW has the best forecast accuracy. The *c* parameter is less than 1, suggesting that the patients are becoming more persistent with time. Furthermore, the polarization index is 0.078, suggesting a quite homogeneous cohort in *p*, the propensity to stop the refill. It should also be noted that if BG were the best model, then its parameters would give rather different insights in terms of the polarization index (which is 0.66). Now we consider the Osteoporosis category. The BG has the best forecast accuracy. The mean of the beta distribution is 0.09, suggesting a low dropout

Table 3: LL, AIC, and BIC for All Datasets. Best Models are Bolded

Dataset	LL			AIC			BIC		
	BG	BdW	LCdW	BG	BdW	LCdW	BG	BdW	LCdW
Hypertension	-1629.0	-1626.6	-1623.5	3261.9	3259.2	**3257.0**	**3271.7**	3273.9	3281.6
Ocular Hypertension	-1557.7	-1541.6	-1536.5	3119.3	3089.3	**3082.9**	3129.2	**3104.0**	3107.4
Statin	-1280.4	-1278.3	-1278.3	2564.7	**2562.6**	2566.7	**2574.5**	2577.4	2591.2
Insulin	-1401.6	-1392.7	-1389.7	2807.2	2791.5	**2789.5**	2817.1	**2806.2**	2814.0
Epilepsy	-1520.3	-1519.3	-1516.8	3044.5	3044.6	**3043.6**	**3054.4**	3059.3	3068.1
RA	-1352.2	-1347.1	-1346.5	2708.4	**2700.2**	2703.0	2718.2	**2714.9**	2727.5
Osteoporosis	-1242.8	-1239.6	-1239.6	2489.5	**2485.3**	2489.1	**2499.3**	2500.0	2513.6
Alzheimer	-1362.4	-1362.3	-1362.0	**2728.8**	2730.7	2734.0	**2738.6**	2745.4	2758.5
ADHD	-970.7	-965.5	-965.4	1945.5	**1937.0**	1940.9	1955.3	**1951.7**	1965.4
Atrial Fibrillation	-1226.0	-1212.6	-1216.0	2456.1	**2431.1**	2442.0	2465.9	**2445.9**	2466.5

propensity in the mean. But its polarization is 0.13, suggesting some polarization in the propensity to stop.

Model Selection

A natural question that might arise is how to choose the best model, among the ones we propose, for a given dataset. Our recommendation is to use out of sample accuracy i.e., accuracy from test data set as a model selection criteria, as we did here. We have particularly shown that one cannot rely on just a *single* model. Other selection criteria such as Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC) can be used as well. AIC and BIC can be easily calculated from equation 9 and 10 respectively, using the log-likelihood (LL) function in equation 7. Since sample size was not used in any of the log-likelihood calculations, we used n = 1000 for all the 10 datasets. k is the number of parameters in the model. Calculated AIC and BIC are shown in Table 3. Best models are highlighted in bold.

(9) $AIC = 2\,k - 2\,LL$

(10) $BIC = log(n)\,k - 2LL$

Table 4 shows the comparison of best models selected based on out of sample MAPE from Table 1 and AIC from Table 3. Between these two model selection strategies, only 4 out of 10 agree. As Armstrong notes, never use in-sample fit for model selection[16]. We agree, and one should always use out of sample data for model selection when feasible. In cases where the sample size is low, AIC/BIC could be used.

Following are 3 simple model selection strategies that can be used in practice:

1. If historical data is available (T >=12 data points), always use out of sample accuracy measures for model selection.
2. If historical data is limited (T < 6 data points) use information criterion such as AIC and BIC.
3. A much better alternative to #2 is to use ensemble models by combining all 3 models by using simple average, median or weighted average of predictions based on prior model performance.

Table 4: Comparison of Models Selected Based on Out of Sample Performance and AIC

	Best Model Based on Out of Sample MAPE	Best Model Based on AIC
Hypertension	LCdW	LCdW
Ocular Hypertension	LCdW	LCdW
Statin	BdW	BdW
Insulin	BG	LCdW
Epilepsy	BG	LCdW
RA	BdW	BdW
Osteoporosis	BG	BdW
Alzheimer	BG	BG
ADHD	BG	BdW
Atrial Fibrillation	BG	BdW

While the analysts would ideally like to know which model would provide the best forecast *before* estimation, this type of work is scant in literature and complex to do (see Schwartz et al.[17]). This question is also beyond the scope of this paper.

Conclusion

In this article we empirically analyzed the effectiveness of probability based model's BG, BdW and LCdW to project patient persistency. We clearly demonstrate that these models significantly outperform excel based trend models. Analysts should consider using these probability based models to project patient retention.

When using these models, we can extract additional diagnostics about the characteristics of the patients' behavior. We already discussed the meaning of the *c* parameter in a discrete-Weibull distribution and there are different forms of heterogeneity. The Beta heterogeneity models, BG and BdW, further allow us to use the parameter values of the Beta distribution to understand the shape of refill propensities across the cohort patients. For an analysis of a specific persistency dataset, these are useful interpretations to know, which are not available from a curve-fitting type regression.

Having an accurate forecast of persistency can allow companies to plan future strategies accordingly. An additional benefit of having an accurate forecast is that it helps to understand the impact of unexpected industry shifts (e.g., new competitor entry, including specialty medicines, better formulary status for most patients, and more convenient mode of administration, etc.). While our models cannot predict what would happen if any of these events occurs (in fact, no model can), at least these models can be used as the benchmark measure for evaluation in an ad-hoc study of event analysis.

One limitation of the current work is that while individual-level data are difficult to obtain, some variables are observable at the patient-level. For example, gender, age, insurance types, etc., are variables that claims databases have. While these variables might not be the causal factors for persistency, they remain of interest to the analyst to understand how they are associated with persistency. Future research will explore the possibility of incorporating these observable covariates in the process of modeling.

In summary, the probability models can provide more accurate forecasts, behavioral insights about the patient cohorts, and computational simplicity. We highly recommend them being the models of choice for projecting patient persistency.

Disclaimer:
The views expressed are those of the authors and do not necessarily represent those of their employers.

About the Authors

Srihari Jaganathan *is head of Advanced Analytics with UCB Inc. Srihari has over 12 years of experience in the pharmaceutical industry. His primary interests are in applying simple analytical and statistical models to drive better decisions in commercial organizations. He has worked across diverse areas within the pharmaceutical industry such as forecasting, health outcomes research, new product launch analytics, business development, sales force effectiveness, resource allocation and optimization.*

Srihari has widely published in various medical and analytics conferences and journals. He has competed in Kaggle Walmart Recruiting - Store Sales Forecasting and was placed 2nd out of 691 competitors from across the world. Srihari recently competed in the M4 time series forecasting competition and was placed 4th out of several competitors in the world. He was awarded the Best Podium Presentation award at the PMSA 2017 Annual Conference. Srihari was also awarded a Gold Medal by the Academy of Managed Care Pharmacy (AMCP) at their 2017 conference.

Ka Lok Lee *is currently a Senior Associate Director, Advanced Analytics and Consulting Services at IQVIA. Ka has over 12 years of experience in applying marketing science techniques to the pharmaceutical industry and has been with IQVIA since 2010. He has published numerous research papers in top-tier academic journals, including Marketing Science and Journal of Marketing Research. His works have been recognized with the Paul E. Green Award from the American Marketing Association, the John D.C. Little Best Paper Award from the INFORMS, the David K. Hardin Memorial Award, and the William F. O'Dell Award for long-term contribution to marketing.*

Ka received his M.S. in Marketing from the Wharton School of the University of Pennsylvania. He also received a B.S. in Economics and a B.A. in Mathematics, both from the University of Pennsylvania.

References

1 Lee K, Fader P, Hardie B, How to Project Patient Persistency. *Foresight: The International Journal of Applied Forecasting*.2007;8:31-35.

2 Fader P, Hardie B, Liu Y, Davin J, Steenburgh T. "How to Project Customer Retention" Revisited: The Role of Duration Dependence. *Journal of Interactive Marketing*. 2018;43:1-16.

3 Kishimoto H, Maehara M. Compliance and persistence with daily, weekly, and monthly bisphosphonates for osteoporosis in Japan: analysis of data from the CISA. *Archives of Osteoporosis*. 2015;10(27):1–6.

4 Fader P, Hardie B. How to project customer retention. *Journal of Interactive Marketing*. 2007;21(1):76-90.

5 Klein J, Moeschberger M. Survival analysis: techniques for censored and truncated data. New York: Springer; 1997.

6 Solomon M, Goldman D, Joyce G, Escarce J. Cost Sharing and the Initiation of Drug Therapy for the Chronically Ill. *Archives of Internal Medicine*. 2009;169(8):740-748.

7 Campbell J, Schwartz G, LaBounty B, Kowalski J, Patel. Patient adherence and persistence with topical ocular hypotensive therapy in real-world practice: a comparison of bimatoprost 0.01% and travoprost Z 0.004% ophthalmic solutions. *Clinical Ophthalmology*. 2014;8:927–935.

8 Kiss Z, Nagy L, Reiber I, Paragh G, Molnar M, Rokszin G et al. Persistence with statin therapy in Hungary. *Archives of Medical Science*. 2013;9(3):409-417.

9 Roussel R, Charbonnel B, Behar M, Gourmelen J, Emery C, Detournay B. Persistence with Insulin Therapy in Patients with Type 2 Diabetes in France: An Insurance Claims Study. *Diabetes Therapy*. 2016;7(3):537-549.

10 Lai E, Hsieh C, Su C, Yang Y, Huang C, Lin S et al. Comparative persistence of antiepileptic drugs in patients with epilepsy: A STROBE-compliant retrospective cohort study. *Medicine*. 2016;95(35):e4481.

11 Neovius M, Arkema E, Olsson H, Eriksson J, Kristensen L, Simard J et al. Drug survival on TNF inhibitors in patients with rheumatoid arthritis comparison of adalimumab, etanercept and infliximab. *Annals of the Rheumatic Diseases*. 2013;74(2):354-360.

12 Suh D, Thomas S, Valiyeva E, Arcona S, Vo L. Drug persistency of two cholinesterase inhibitors: rivastigmine versus donepezil in elderly patients with Alzheimer's disease. *Drugs & Aging*. 2005;22(8):695-707.

13 Beau-Lejdstrom R, Douglas I, Evans S, Smeeth L. Latest trends in ADHD drug prescribing patterns in children in the UK: prevalence, incidence and persistence. *BMJ Open*. 2016;6(6):1–8.

14 Gomes T, Mamdani M, Holbrook A, Paterson J, Juurlink D. Persistence With Therapy Among Patients Treated With Warfarin for Atrial Fibrillation. Archives of Internal Medicine. 2012;172(21):1687-1689.

15 Akima H, Gebhardt A. akima: Interpolation of Irregularly and Regularly Spaced Data. R package version 0.6-2. Available from: https://CRAN.R-project.org/package=akima [Accessed 15th October 2018].

16 Armstrong JS, editor. Principles of forecasting: a handbook for researchers and practitioners. *Springer Science & Business Media*; 2001 May 31.

17 Schwartz E, Bradlow E, Fader P. Model Selection Using Database Characteristics: Developing a Classification Tree for Longitudinal Incidence Data. *Marketing Science*. 2014;33(2):188-205.

An Approach to Operationalizing Next Best Action in Pharmaceutical Communications and Marketing

Marc-david Cohen, Ph.D., Chief Science Officer, Aktana Inc.; Amy Baer, Senior Data Scientist, Aktana Inc.; Michael Steiner, Senior Data Scientist, Aktana Inc.

Abstract: The growth of on-line commerce has focused a lot of attention on finding algorithms and approaches to identifying the next best action (NBA) in personalized marketing and sales communications. When the Netflix challenge was launched in 2006 it started a wave of research into collaborative filtering and other approaches to recommendations. The algorithmic approaches that emerged from this challenge continue to be used in the quest to find the personalized message, its timing, the best delivery channel, and message content that is most impactful to the bottom line. However, there is a big difference in complexity between a recommendation for purchase in a retail setting and leading a health care provider (HCP) on a journey to improve pharmaceutical sales and brand loyalty.

Some of this complexity stems from the large number of decision variables, including action timing, and covariates that must be accounted for in an all-inclusive model. Because of the high dimensionality of this problem a large number of observations would be needed to build an accurate and predictive model across the entire space. Then, to use such a predictive model in practice the impact of counterfactuals of the decision variables would have to be estimated and evaluated. The high dimensionality would mean that a very large space would have to be searched to find the estimated best action.

Unfortunately, for model training real-world sales operations data are often limited, so it is likely that large segments of the decision space are not adequately represented in the training data. We propose decomposing the decision space, training separate models for different decisions, and then finding the decisions that maximize these marginal models separately and independently. Strictly speaking the resulting solution will be sub-optimal since it discounts covariance between decisions determined from these independently trained models. However, if there is little training data that would be available for the single all-encompassing model one might expect that the loss from suboptimal decisions would be minimal.

For simplicity we motivate these ideas with an example that has the objective of maximizing email opens. We discuss how this might be applied with several conditional models including ones that (1) predict next message to send, (2) timing of the message send, and (3) wording of the email or message topic.

We use a simulation to show the tradeoffs of using conditional models. The simulation, as in practice, uses a non-linear machine learning approach to train a logistic function with the objective of being able to predict which of a set of possible actions maximizes a zero-one target such as email opens.

Finally, there is additional value to training and integrating separate models into a large automated production environment by minimizing change in the execution environment. This means that

- Models are quicker to implement, require less data, and yet can approximate the NBA;
- Innovation can be brought to the client much more rapidly with minimal if any cost in predictive strength from using component models;
- Additional data can be more easily brought into existing models to improve predictive accuracy on component targets.

Keywords: Next best action, Prediction accuracy, Machine learning, Modeling covariance, Conditional models, Real-world data

Introduction

There has been a lot of discussion in the popular press about next best action (NBA). This concept evolved from market-basket analyses to identify product affinities – if you bought this product then you'd be likely to buy this other product. Much has been written about this and a lot of analytic methodologies have been developed and refined to solve this problem. The Netflix challenge in the early 2000s focused attention on the problem and resulted in significant improvements in predictive analytics, methodologies, and general approaches to predicting what people "like" or are interested in given their history of affinities and purchases.

Much of this work evolved into "recommendation engines" and other concepts that would recommend to consumers a product or service that they would likely be interested in given their history of purchases. Another concept that evolved as part of this discussion and during this period is the idea of 1-1 marketing or personalized marketing. This approach seeks to present offers to consumers that are individually tailored to them based on their historical purchasing and interest patterns.

Recently, these concepts and other related ideas have been developing within pharmaceutical brand management and sales operations with an effort to market to physicians and other health care providers (HCP) in a personalized way, namely personalized to the individual HCP, so that contact strategies and specific communications are made to the HCPs based on the history of communications and interactions they have had and on how the HCPs reacted to those encounters. If one considers all the possible types of communications (and actions such as sampling) and channels for delivery, the combinatorics of the decision problem becomes very large, and simply identifying the next best communication to deliver and over which channel to deliver that communication is difficult.

Ultimately, the purpose of all the communications is to increase sales, prescription writing, and use of the therapeutic. The goal of NBA is to inform this communication strategy and lead the HCP on a journey to optimize the sales objective. However, the scale of decision making on individual communication's impact on sales is difficult, if not impossible, to measure. So, other measures of performance for communications should be considered. Things like did the HCP read an email sent; did the HCP seek more information after a face to face visit by a sales rep; did the HCP go to a website after opening an email; and many other indicators of HCP engagement with the brand.

Consider a framework to NBA where there are two broad components to the decision problem. The first component is an estimate of a response function that captures the relationship between all the relevant variables and the measure of performance. Relevant variables include all we know about the HCP and their practice and the history of communications to the HCP, and whether the performance measure is either discrete or continuous. The second component is a decision model based on the response function. Some of the dimensions of the response function can be considered decision variables. These are the variables under control. For a given observation of an HCP some of the dimensions are fixed and known and others, the dimensions associated with the decision variables, are the ones used to maximize the response function.

One big caveat to this approach is the assumption of causality. In some domains it is

very important to identify causality as distinct from correlation. That is certainly true in this domain of marketing and is a focus of effort when making decisions at a segment level that are associated with significant budgets. However, we will sidestep that discussion in our approach and instead focus on our ability to leverage the historical data with automated modeling, associations, and continuous learning.

Next Best Action

Predicting the next best action (NBA) is a formidable task if we consider at any point in time all the possible actions that can be taken that might positively affect the outcome. For simplicity we motivate these ideas with an example that has the objective of maximizing email opens. Consider alternative actions that can be taken at the HCP level: Visit, Send an email, Send a letter, Provide a sample, Invite the HCP to a seminar, Invite the HCP to a webinar, Take any of these actions now or delay, Select the content of the discussion, email, letter, or webinar, Wording for the email header, etc.

Consider how the impact of these choices might be measured, perhaps by some immediate feedback such as whether the HCP opens the email or clicks on a link within the email, or accepts the invitation to the seminar or webinar, or provides direct feedback from a query during the visit. Each of these responses are closely tied to the alternative actions. Measures of script writing as a response are not as closely tied to the direct action, they are more cumulative in nature as a result of many interactions and are much harder to tie to any specific action. Here we focus on the direct feedback resulting from an action since our attention is on the direct 1-1 impact on the HCP.

Let's consider the target that we want to impact is whether the HCP opens an email. The decision variables for this might be:

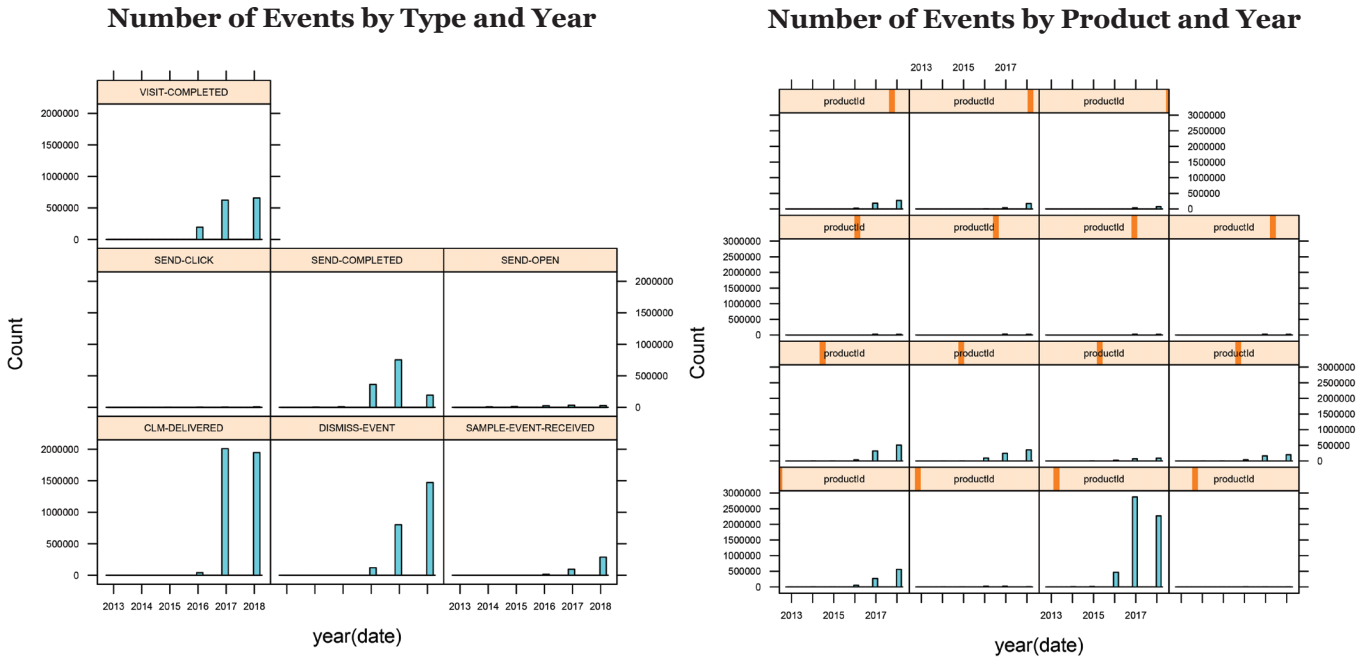
- the timing of email,
- time of day email is sent,
- day of the week email is sent,
- when the last visit occurred,
- the content of the email as indicated in the email header,
- the wording of the email header.

For example, we may ask, "Do certain email headers lead to increased open rates for certain HCPs?". We can then tailor the decision variable accordingly by suggesting reps use these headers when sending emails to those HCPs. In addition to decision variables as predictors other variables are useful such as, the history of contacts to the HCP that included the combinations of values for these decision variables as well as covariates for the HCPs that include HCP demographic information, data on the practice, HCP preferences (say for channels and therapeutics), 360 views of their communication, and if possible insights into HCP patient population and other covariates that have been collected. Provided with enough data a single model with all these decision variables might be built and used for prediction and making decisions. It quickly becomes apparent that this is a very large decision space and would require a significant data set to adequately fit a response function which could be used to estimate.

Of course, the general goal is to increase script writing and sales performance. We assume that the actions that result in more immediate feedback, such as email opens, as a measure of HCP interest and engagement can cumulatively result in improved sales performance.

There are many decisions such as those enumerated above that impact email open likelihood. In the simulation discussion below, we explore the impact of developing separate response models for each of these decisions and then using a product of the separate response models in the decision model.

Figure 1: Histograms of Real-World Data



Note that many of the ideas developed here also apply to the goal of increased sales and as well can be applied when sales is a target of decision making. These will not be explored in this paper. The next section overviews some real-world data that illustrates the size and scope of this data issue.

Typical Real-World Data

A concrete example based on data we have analyzed can provide more clarity on the problem, the dimensionality, and complexity. Consider a single brand that has about 650K HCP accounts. Let's assume that for this brand and the desired decision model, the performance is measured by an aggregate of HCP outbound communication. For this brand there are three channels, email, snail mail, and face to face meetings with the following dimensions:

- 15 products
- 90 potential email templates covering various brand related topics
- 65 physical letters covering various brand related topics
- 12 topics for delivery directly through face to face rep meetings

Each contact with an HCP can be characterized, or described, by these various dimensions. If these were all mutually exclusive in describing the contact, there would be over 1 million possible combinations. Although that is not the case since many do not make sense together, for example sending an email and a physical letter in the same contact, there are many combinations that are legitimate.

Between October 2013 and October 2018 there were approximately 30 million contacts between the approximately 3000 sales reps and the 650K HCPs. This averages out to about 3 contacts per month for each HCP. If we look at summaries of a few of these dimensions by year we can get a sense of the concentration of the data. The histograms in Figure 1 show the distribution of event categories by time and product. Note that each communication contains a message delivered in that communication.

In addition to the sparsity in the contact history each HCP may be characterized by various variables and grouped by segment. Often brand

Table 1

Variable	% Missing	Distinct Values
Brand Segment	43.7%	18
Digital Preference	95.5%	2
Therapeutic Preference	91.2%	58
Therapeutic Segment	98.3%	5
Therapeutic RX Quintile	98.2%	6
Multiple Therapeutic Targets	0%	2
Brand Priority	0%	7

teams do deep analysis and use the results of that analysis to segment and control the contact strategy. While our ultimate goal centers on personalization, segmentation can be used for modeling in the event that we do not have enough historic data and due to sparsity. In our example data there are 45 HCP segment and characteristic variables. Table 1 shows a select few of these variables and their relative density within the dataset. Notice that for some of the variables there are a large number of missing values.

If we look across all these account variables there are more than 147K unique combinations. This means that on average there are 4-5 HCPs out of the 650K HCPs that share the same values for these 45 HCP segment and characteristic variables.

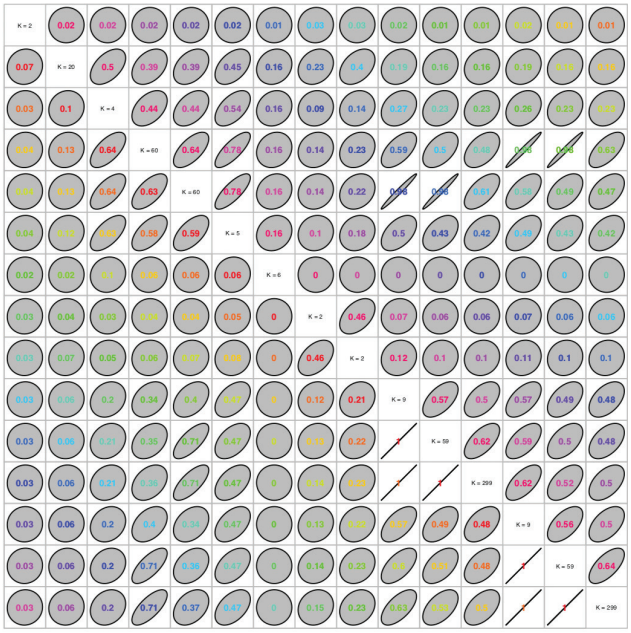
In addition to the channels and the HCP characteristics there are various timings over which communications can be delivered. For example, NBA not only includes the message and channel dimensions but also includes the time dimension. For the complete solution we want to consider the effect of various timings on delivering the message. For example, we would want to account for first visiting an HCP and then sending an email in comparison to first sending an email and then visiting. And, the timing between the visit and the send could vary based on contact history and the nature of the visit and the email.

The following plot shows some of the Goodman Kruskal correlations between some of the HCP characteristics. To protect the client's data, we've removed the names of the selected variables from the graph. Typically, these variables describe the HCP, like gender, age, geographic location, regional population where HCP practices, HCP's average market share for the target therapeutic, educational history, publication history, and other potentially descriptive characteristics. Unlike the correlation coefficient this statistic is not symmetric so that it captures some directionality in the correlations. Notice that there are pockets of higher correlations and that many of the variables have few values for multiple HCPs. The number of HCPs with multiple values for two variables is shown in the diagonal. The simulation that we report on following will attempt to capture some of this structure.

Simulation

As mentioned above the general approach to the decision problem is to model a response surface from the complete set of predictors to a target that is desirable to optimize. We do not focus on the nature of the target variable. Suffice it to say that it is some measure of HCP satisfaction that is highly correlated to client revenue for the target therapeutic. We assume that it is correlated to the collection of predictors. We will also not address the issue of causality vs correlation in the predictors. Although both of

Figure 2: Correlations of Real-World Data



these topics are very important, they are not the focus of this paper and would require a much more extensive research. We and our clients assume that some set of predictors is causal and drives the target performance.

The general approach is to fit a response curve to the target. Then, based on the client understanding of causality a set of variables is identified as the decision variables. The resulting decision problem is to find the values of the decision variables that optimize the target for each observation. We fit the response curve using machine learning algorithms such as random forests, gbm, or other technique and then for each HCP fix their characteristic variables and history, and predict the response using the modeled response curve as we vary the decision variables across the range of actions for them.

The core question we explore is the cost of ignoring correlation when building a response surface model. This approach can result in a large prediction model for the response surface.

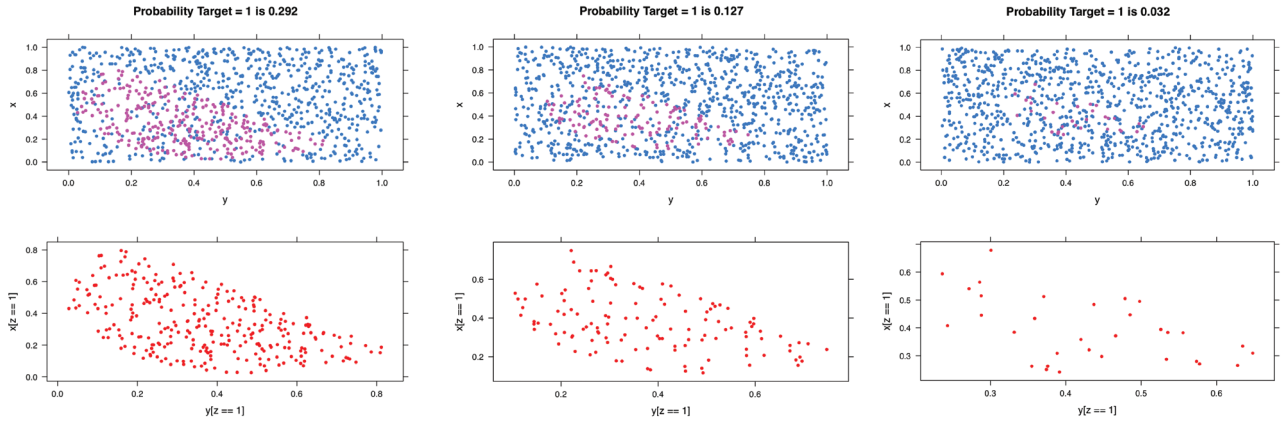
Let X, Y be a set of predictors and Zε{0,1} be a binary target. Then, the response surface is:

$$pr(Z|X,Y) \approx pr(Z|X)pr(Z|Y)$$

Simulation Setup

We explore this with a simulation that replicates some of the structure that we observe in real-world data; namely the sparsity and the small response likelihood in a binary setting. We will explore this across three dimensions, (X,Y,Z), where two of the dimensions (X,Y) are continuous and uncorrelated with each other and the third Z is discrete [0,1] and correlated with each of the other dimensions with correlation ranging from approximately -.08 to -.25. An example of the relationship between this simulation setup and the problem as discussed above is to consider one of the continuous dimensions, say X, as an HCP characteristic and the other, say Y, as a decision variable, for example, what time to send an email. Then, the variable Z is the outcome as to whether the email is opened or not. In practice

Figure 3: Three Samples



both X and Y would be multidimensional even though Z would be binary or continuous.

We build three response models each a logistic function estimated using the gradient boosted models with the gbm() function in R.

- a Joint model predicting pr(Z|X,Y),
- an X (marginal) model predicting pr(Z|X), and
- a Y (marginal) model predicting pr(Z|Y).

We then analyze the decision model comparing the performance of the models of pr(Z|X,Y) to pr(Z|X)pr(Z|Y) for various cutoffs for the decision Z = 1 and how the correlation ignoring strategy compares.

The simulation is a grossly simplified 2-dimensional model in contrast to the real-world scenarios where the model space can be 100s of dimensions and much more sparse. We have attempted to emulate the sparsity of the actual responses being fit. The 2 dimensions are both uniform [0,1] random variables with 1000 sampled points. The responses are modeled by a bivariate normal random variable with mean (.5,.5) and variance matrix $\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. We set the target to one for points in the lower left diagonal of the sampled hypercube ($X+Y<1$) and for values where the sampled bivariate Normal is greater than a specified quantile of the Normal sample. A separate simulation is run and analyzed for each quantile from .1 to 1.

The three plots in the first row (Figure 3) show the samples for the .25, .5, and .75 quantiles. Note that the titles of the plots show actual probability of open, namely Σ(Target = 1)/N where N = 1000 points in the sampled hypercube. The points in red are the observations where the target is 1 and the title shows the actual probability that the target is 1 for each of the samples. The second row just shows the observations where the target is 1.

Simulation Process

The execution logic for each of the 5 simulations is as follows:

- Sample the predictor hypercube and the target=1 points.
- Use Generalized Gradient Boosting (gbm) for fitting three logistic regression models: pr(Z|X,Y), joint model with predictors x and y; pr(Z|X), a marginal model with predictor x; and pr(Z|Y), a marginal model with predictor y.
- Produce various plots to compare the joint, marginal, and product of marginal prediction models fit

Figure 4: Predictions for the Joint Model and X Marginal Model

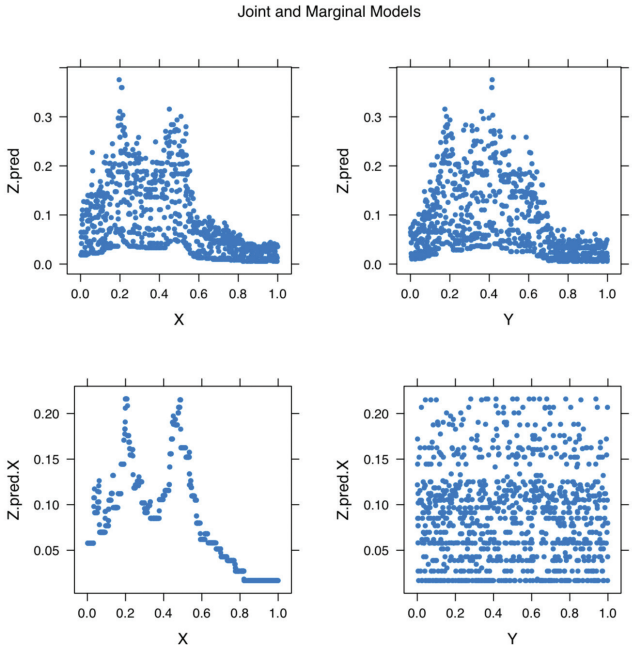
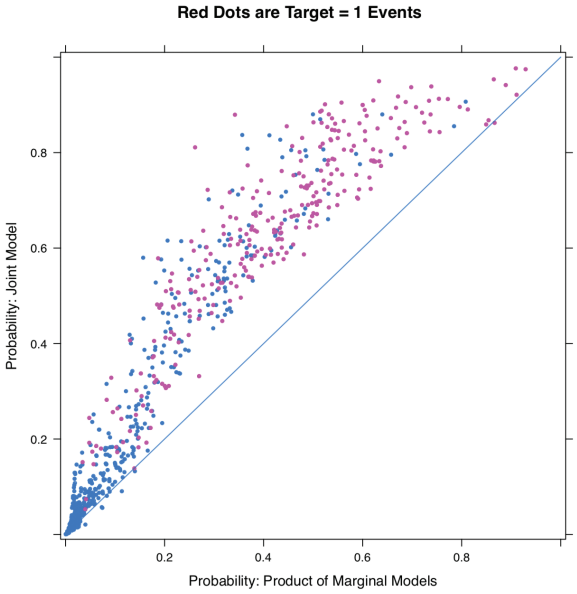


Figure 5: Joint Predictions pr(Z|X,Y) vs the Product of Marginal Predictions pr(Z|X)pr(Z|Y)



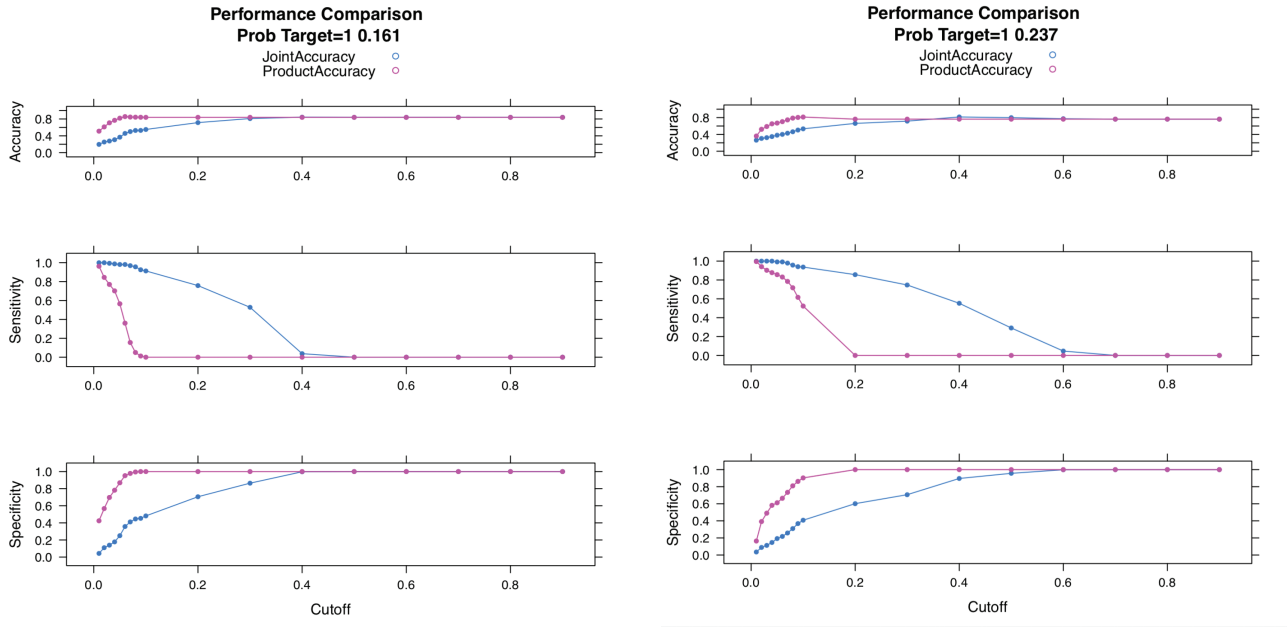
- Compute the confusion matrix for the joint and product models for various cutoffs that might be used in the decision model.
- Produce various plots to compare the decision model performance for the joint and product of marginal prediction models.

The plots in Figure 4 show the predictions for 2 logistic models estimated by the gbm function on the data in the left column in Figure 3 with probability of target=1 of .292. The first row shows the results for the joint model where both x and y were used to estimate the probability that the target is 1. The left plot on the first row shows the joint prediction as a function of the x dimension and the right plot shows the prediction from the joint model as a function of the y dimension. It is interesting to compare the x marginal model predictions as a function of x (lower left plot) vs the prediction as a function of y (lower right plot). As expected, the estimates of the probability of target=1 for the x marginal model show no structure when plotted as a function of y.

Figure 5 is a plot of pr(Z|X,Y) vs pr(Z|X)*pr(Z|Y). If there were no correlation between the Z and X, Y we would expect the points to line up on the diagonal because in that case pr(Z|X,Y) = pr(Z|X)pr(Z|Y). The plot shows points where Z = 1 in red. The decision model is derived from the prediction model by providing a probability cutoff value above which the decision is to send the email with the presumption that the email will be opened with probability derived from the prediction model. So, one way to compare the performance of the decision model across the two methods of prediction, one would choose a cutoff value, say .5, and then look at the statistics of performance for the Z=1 vs Z=0 for all observations above that cutoff in each dimension.

One tool commonly used is the confusion matrix and the statistics based on it such as accuracy, sensitivity, and specificity.

Figure 6: Comparison of Decision Model Accuracy, Sensitivity, and Specificity



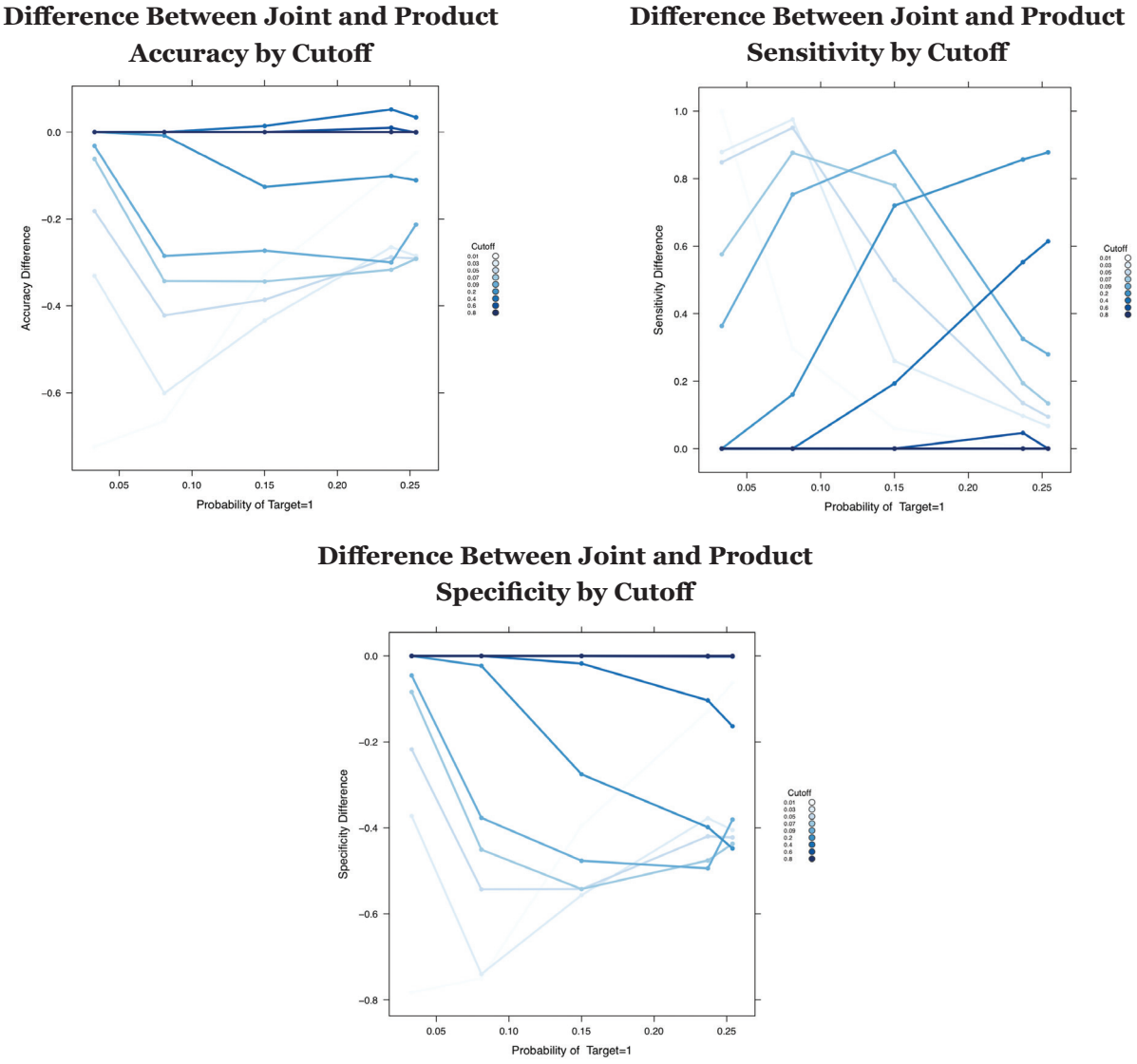
Simulation Results

Figure 6 shows such a comparison for two of the simulations. The plots show a comparison of the joint estimates and the product estimates across a spectrum of cutoff probabilities. Observations from the plots in this figure are consistent with what is observed across the simulations which are not shown here.

- The Accuracy of the product estimates is consistently better than that for the joint estimates. The Accuracy measure is the ratio of the sum of the predicted positives that are correctly identified as positive (Sensitivity) and the predicted negatives that are correctly identified as negative (Specificity) divided by the total population. As a result, it contains two sources of error - the sensitivity and the specificity. These represent type I and type II errors.
- The Sensitivity of the product estimates are consistently below that of the joint estimates.
- The Specificity of the product estimates are consistently above that of the joint estimates.
- The differences in these measures converge as the cutoff grows to include larger proportions of the population.

These plots of accuracy, sensitivity, and specificity show that accuracy of the product estimates is higher than the joint estimates because their sensitivity is much inferior to the joint estimates. This means that a decision model based on the product estimates is likely to take action more aggressively and make many more type I errors and many fewer type II errors. So, if the decision is whether to send an email the product estimate will send more emails to people who are not going to open them than the joint estimate base decision model would. But the product model would be less likely not to send email to people who would open them than the joint model. In the context of a marketing decision this is probably desirable since the cost of over-reaching is most likely lower than the cost of under-reaching—which in the context of sending email might be spamming the target audience. If it is more important to avoid spamming, then the methodology could be applied but with reversing the interpretation of the target.

Figure 7: Overall Accuracy, Sensitivity, Specificity by Cutoff



One final note concerns the choice of cutoff value. This is typically done by optimizing a loss function of the specificity and sensitivity that captures the costs associated with the different kinds of errors. The optimum of this may not be the same as the accuracy maximizing point.

The plots in Figure 7 show the differences between the performance of the joint and product estimates across the range of cutoffs for all the probability of Target=1 choices in the simulation and for the three performance measures. For each of these plots, if the difference is positive the joint estimate is "better" than the product estimate and vice versa.

The leftmost plot shows the differences in the accuracy measure as a function of the target probability in the raw data, in other words, if the probability of Target=1 is .15 that means that 15% of the population has a target of one. The various differences are for each of the various cutoffs tested across all the target probabilities. As one would expect, as the cutoff

increases to include the entire population (cutoff=1) the differences across each of the measures converge. It is interesting to note that the accuracy and specificity differences are almost monotonically changing with the cutoff value but that the sensitivity shows a change in direction as cutoffs increase.

Conclusion

Operationalizing NBA decisions is a complex issue and involves many tradeoffs. This paper attempts to focus on one small compromise that we believe makes it a more realistic and reachable goal and provides the beginnings of a divide and conquer methodology. We feel this approach helps to add NBA to an automated system designed to provide recommendations and suggestions on a real-time basis in an operational setting. Our experience is that sales operations are very sensitive to changes in an operational system and what impacts those changes might have on suggestions. In practice, if we can limit changes it will limit the cost of putting those changes in place.

This simulation just begins to explore the notion that one can use pr(Z|X)pr(Z|Y) for pr(Z|X,Y) and ignore known correlations. However, we feel there are some explorations that would help to illuminate the tradeoff that this compromise makes. These include:

- Replications - doing independent replications of this and like-minded simulations to understand the variance associated with these results.
- Higher dimensions - making the predictors of higher dimensions so that greater insight into the effect of sparsity on the compromise could be obtained.
- Correlation among some of the predictors - the 2-dimensional simulation we performed were independent. This avoided multicollinearity which might have an impact on the machine learning based logistic models that resulted.
- More advanced functions characterizing the target - what would the impact be if instead of a binary outcome variable the outcome was ordinal or continuous.

The simulation and much of the discussion does not address the fact that for NBA we are interested in multiple responses that impact overall HCP value. This is a large topic and could not be addressed in a short paper. However, the concepts developed here can be generalized to predict a single measure of performance across the multiple predictive dimensions.

About the Authors

***Marc-david Cohen**, Ph.D., Chief Science Officer, Aktana, is an experienced business leader with a background in operations research and statistics. At Aktana he leads the development of learning and insight generation capabilities. Previously he served as CSO at Archimedes Inc. - a Kaiser Permanente Company - and helped the company transform from HEOR pharmaceutical consulting to a products company focused on clinical studies and personalized medicine. Previous roles included VP of Research at FICO and multiple senior roles at SAS Institute where he initiated the SAS Marketing Optimization product.*

***Amy Baer**, Senior Data Scientist, Aktana, is a biostatistician with a background in predictive analytics. At Aktana she conducts research-related projects for the machine learning team including improvement of model performance and feature interpretability. Previously, she worked at MedAmerica as a statistician focusing on preventative patient care and emergency room operations. Amy holds a master's degree in biostatistics from California State University, East Bay and a bachelor's degree in mathematics from The University of Chicago.*

***Michael Steiner**, Senior Data Scientist, Aktana, is a mathematician with a background in topological data analysis. At Aktana he builds prototype models in order to improve and extend the suite of services offered, and he works on special projects for clients in a consulting role. Previously, he worked at Forio as a Data Scientist focusing on simulations. Michael holds a master's degree in mathematics from San Francisco State University.*

References

1 Altmueller, S., Haralick, R., "Approximating High Dimensional Probability Distributions," Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), pp. 299-302, Aug 2004.

2 Chessell, M., Pugh, D., "Smarter Analytics: Driving Customer Interactions with the IBM Next Best Action Solution," http://www.redbooks.ibm.com/abstracts/redp4888.html?Open. Aug 2013.

3 Generalized Boosted Regression Models R package. https://cran.r-project.org/web/packages/gbm/gbm.pdf.

4 Islam, M., Chowdhury, R., Briollais, L., "A Bivariate Binary Model for Testing Dependence in Outcomes," *Bulletin of the Malaysian Mathematical Sciences Society*, Vol. 35 Issue 4, pp. 845-858, 2012.

5 Lee, Y., Nelder, J., "Conditional and Marginal Models: Another View," *Statistical Science*, pp. 219-238, 2004.

6 Li T., Zhu S., Ogihara M., Cheng Y., "Estimating Joint Probabilities from Marginal Ones". In: Kambayashi Y., Winiwarter W., Arikawa M. (eds) Data Warehousing and Knowledge Discovery. DaWaK 2002. *Lecture Notes in Computer Science*, Vol. 2454. Springer, Berlin, Heidelberg

7 Moore, C., "Pharma Marketing: Get Started on Creating Great Customer Experiences with Journey Strategies," https://a.sfdcstatic.com/content/dam/www/ocms/assets/pdf/datasheets/pharma-insight-paper.pdf.

8 Muff, S., Held, L., Keller, L., "Marginal or conditional regression models for correlated non-normal data," *Methods in Ecology and Evolution*, Aug. 2016.

Aligning Product Forecasts to Physician Target Universe

Ester Oben Etah, Principal, Launch Excellence, IQVIA

Abstract: Typically, product forecasts are generated by the forecasting and market analytics teams who are usually looking at market need and product differentiation. Although other aspects such as pricing, manage care restrictions, etc., may be considered, the forecast team usually computes the forecast independent of the brand and marketing team who are responsible for delivering on the forecast.

On the other hand, the brand and marketing team is usually concerned about the level of investment needed to support the product in the market. Their concerns typically revolve around new patient acquisitions and physician initiators. This article, based on analyses of over 500 launch brands, will demonstrate that *new patient acquisition* is a leading indicator of total prescription volume TRx.

Keywords: Launch forecast, Product forecast, TRx forecast, New patient acquisition, NBRx

Background

Bringing in new patients and prescribers and maintaining them have proven to be predictors of the success–or failure–of new pharmaceutical product launches. To excel at launches, particularly within the first critical six months, launch teams must ensure that there is alignment between the forecast and the number of prescribers to be targeted. This is critical in ensuring that product forecasts are both realistic and achievable. Typically, product forecasts are generated by the forecasting and market analytics teams who are usually looking at the following two aspects of a product:

1. **Market Need:** The level to which current drugs and the standard of care is helping patients achieve their therapeutic goals which may be a cure or reaching an acceptable therapeutic target. In this area they will consider current and future events that will drive growth within their specific therapeutic market, such as a new diagnostic technique that will better identify patients in need of therapy or an industry wide event such as regulatory requirements from the FDA or any other key industry stakeholder.
2. **Product Differentiation:** What are the unique characteristics of the product that will enable it to compete effectively in the market? This will address aspects specific to the product such as efficacy, tolerability, convenience, etc., with the goal of determining what section of the market the product will play in and what proportion of that market it will carve out for itself. Their performance will be compared to current products in the market as well as future launches which will have a direct or in-direct impact on the product.

Although other aspects such as pricing, manage care restrictions, etc., may be considered, the forecast team usually computes the forecast independent of the brand and marketing team who are responsible for delivering on the forecast.

On the other hand, the brand and marketing team is usually concerned about the level of

Figure 1: The New-to-Brand (NBRx) – Patient Acquisition Metric

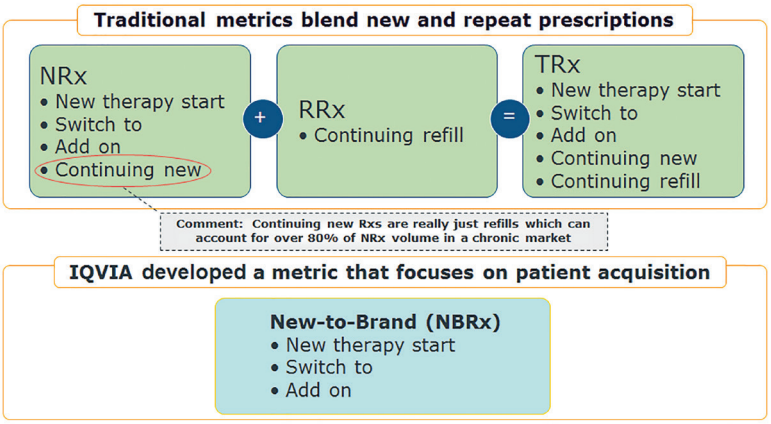
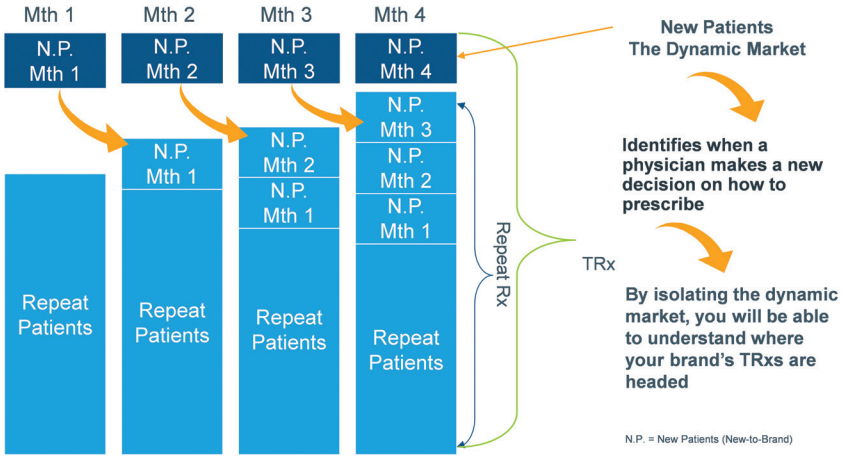


Figure 2: The NBRx to TRx Model



investment needed to support the product in the market. Their concerns typically revolve around:

1. **New Patient Acquisitions:** Of the treatable patient pool, how many of these patients can they actually get to try their product? These may be newly diagnosed, treatment naive patients, patients who are unsatisfied with their current therapy and need something else either to switch products or add the new product for increased efficacy. There is a significant cost associated in acquiring new patients through DTC in (TV, radio or print), digital and social media or through patient advocacy groups.

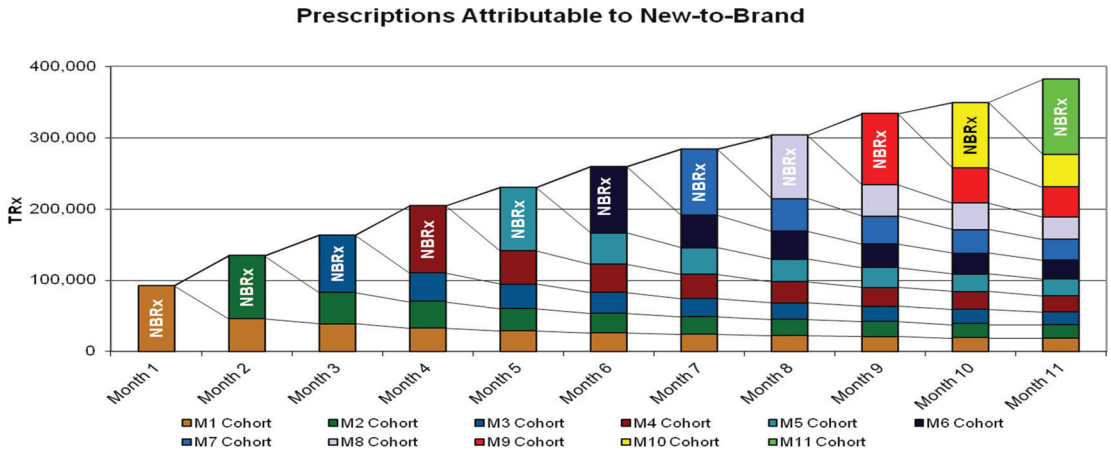
2. **Physician Initiation (New Writers):** Getting healthcare providers to adopt their brand is also very costly. They can also reach physician and other HCPs through DTC, digital, and other channels such as conferences etc., but the most effective way of acquiring new physicians is by direct and personal engagement using sales reps.

We have shown through the analyses of over 500 launch brands that new patient acquisition is a leading indicator of total prescription volume TRx.

New patient acquisition or NBRx (Figure 1) represents distinct medical decisions by the physicians. It is when a physician is deciding to put a patient for the first time on a new medication. Once already on the medication, if they derive some benefit, they will continue to refill and drive TRx volume as shown in Figure 2.

Figure 3

Figure 4: Decomposition of TRxs into NBRxs and Refills for Each Time Period



The NBRx Model (New Patient Acquisition Analysis)

The NBRx model is based on a very simple algebraic concept, shown in Figure 3.

Using this formula, we can decompose the total TRx prescriptions for each time period into NBRxs and Refills (repeat prescriptions) as shown in Figure 4. Refills are generated using refill rates derived from analogues for pre-launched products or from the brand performance for marketed products.

Using this methodology we are able to determine the number of new patients (NBRx) that will be required to achieve the TRx forecast.

The NBRx is a key performance indicator (KPI) because both the brand team and the forecasting team can quickly determine if the NBRx volume required is achievable and whether they can hit the forecast, or if there is a gap, they can determine how big the gap is.

Another KPI that is typically used by the brand team and can also provide useful insights to the forecasting team is the number of physician initiators required to drive the new patient acquisition necessary to hit the forecast. Physician initiators represent the number of new writers who decide to put their patients on the brand.

The Physician Model (Physician Initiation Analysis)

We have also determined that about 60% of new patient acquisition (NBRx volume) is accounted for by physician initiators as shown in Figure 5.

Physician initiation is a big part of the equation, but how those prescribers act in terms of the volume (productivity) and the frequency (persistency) of writing after initially prescribing the product is also a large driver of the success or failure of a launch brand.

Figure 5: The Relationship Between Physician NBRx Initiation and NBRx Volume

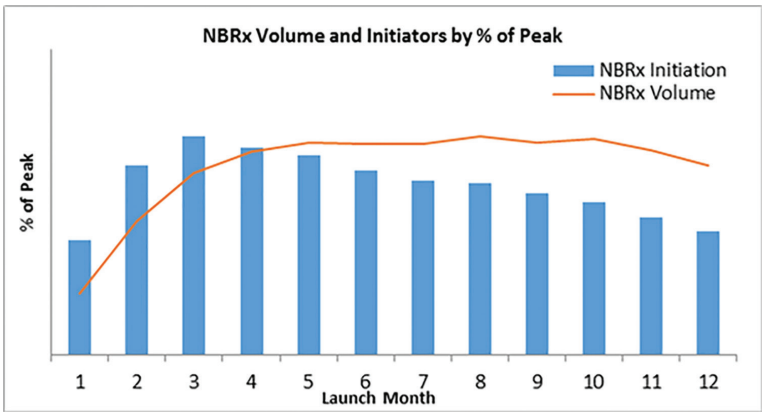
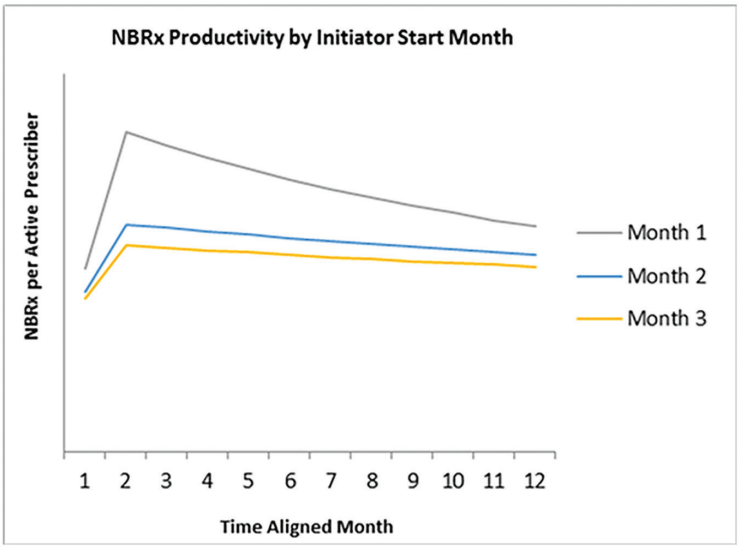


Figure 6: Productivity of Physician Initiators Over Time from Month of Initiation



Productivity determines the volume potential for a launch brand and, therefore, the level of NBRx initiators required, as illustrated in Figure 6.

High physician productivity means fewer initiators are needed, and the early initiators or early adaptors typically have higher productivity compared to physicians initiating later.

Persistency rates, on the other hand, influence the shape of the required NBRx Initiator curve, as illustrated in Figure 7.

Once prescribers have decided to try a product for a patient, either being new to therapy or switching from a prior market therapy, what happens next will depend on several important factors including:

- Available patient base
- Actual/perceived efficacy of the product
- Negative experience with the product, i.e. safety issues or payer access challenges?

The number of times the prescribers write after the initial trial is their persistency rate; low persistency means a product needs to continue to drive initiation in order to see growth in NBRx and vice versa volume.

For a pre-launch brand, we use analogues to estimate the productivity and the frequency of

Figure 7: Persistency of Physician Initiators Over Time from Month of Initiation

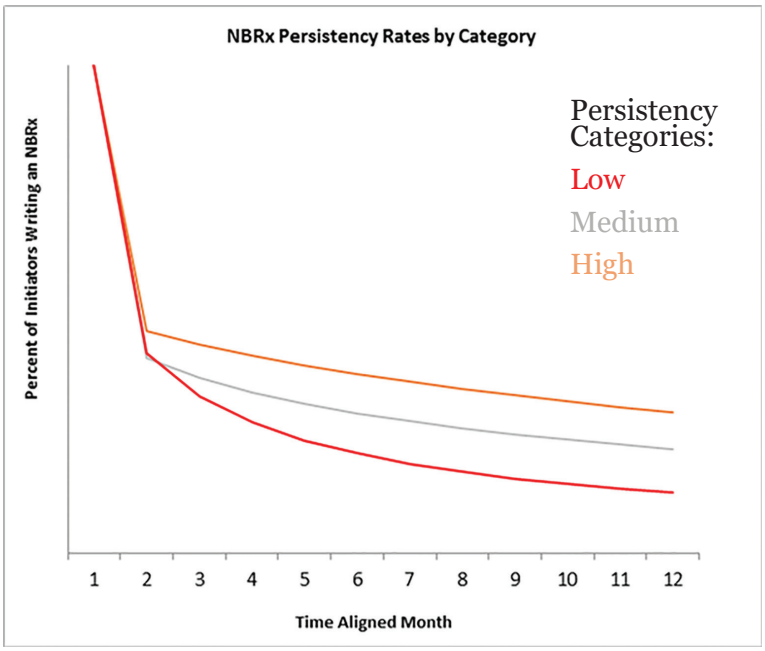


Figure 8

writing for the physicians within the market once they have initiated. For marketed products, we use their historic values.

The Physician model is based on a very simple algebraic concept; see Figure 8. Therefore, Physician Initiators = NBRxs / Physician Value.

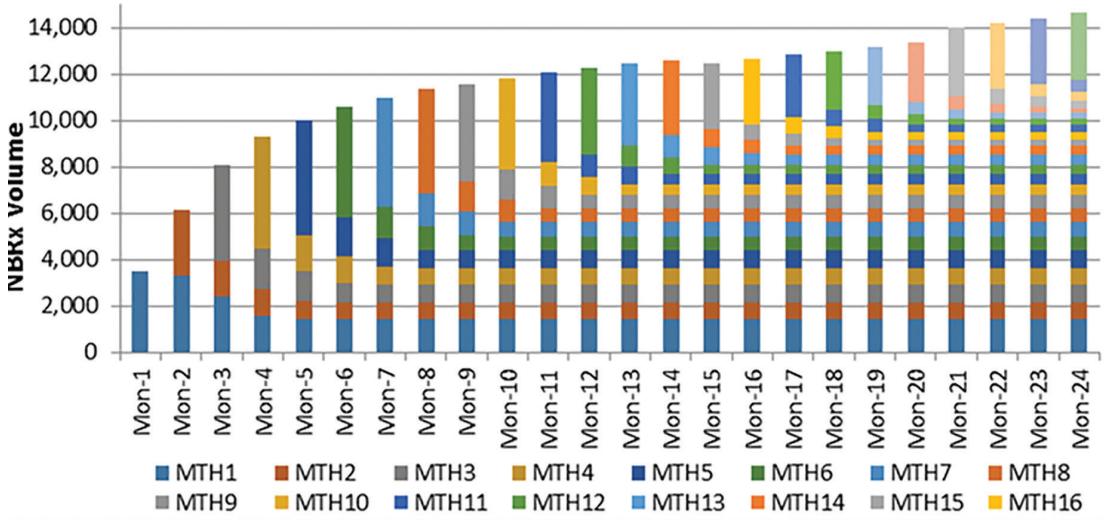
Initiators become repeaters and their NBRx volume is added to initiator volume as shown in Figure 9.

The Physician Model is used to determine the number of Physician Initiators needed to drive the NBRx volumes required to hit the TRx forecast. It is another KPI that can be used to determine if the TRx forecast is realistic or achievable. If there is a big difference between the number of Physician Initiators the forecast is calling for and the universe of target doctors, then the brand team and forecasters need to consider making adjustments.

The NBRx Dynamic Simulator seamlessly brings together both the NBRx and Physician models. Brand teams are using the simulator to answer the following questions:

- **Is my forecast achievable**?
  - Is my preference share achievable?
  - How long will it take to achieve peak share?

Figure 9: NBRxs by Physician Initiation Cohort



  - Is the shape of my uptake curve realistic?
- **Does the size of my salesforce support our forecast?**
  - How many new patients are needed from each prescriber?
  - Is the number of physicians I am targeting enough?
- **How can launch performance be evaluated early?**
  - How many new patients do I need on a monthly or weekly basis to achieve my forecast?
  - How many prescribers do I need to initiate on a monthly basis?
- **Are we still on track to hit our long-term forecast?**
  - NBRx required and TRx forecast validation for in-line brands
  - TRx and sales revenue gap analysis

Below are two case studies that examine this process of alignment utilizing the dynamic simulator tool and NBRx model. In each case, the Launch Excellence team uncovered an issue with the client's forecast. The following describes how the team solved those issues.

Case Study 1

The Launch Excellence team's initial analysis of the company's TRx forecast revealed that the company would need to double the number of physician initiators than they had in their entire target universe, which meant that the number of new patients required as well as the TRx forecast were too aggressive, unrealistic and unattainable.

Method

Using the NBRx Dynamic Simulator, the Launch Excellence team performed the following analyses:

1. **Calculated the New Patients (NBRx) Required**
   Starting with the client's physician target universe, the Launch Excellence team calculated the number of new patients that could be acquired through these physicians based on their productivity, persistency and rate of initiation derived from analogues.
   - NBRxs = Physician Initiators x Physician Value

2. **Calculated the Monthly TRx Forecast**
   Using the modeled NBRxs from step 1, the Launch Excellence team determined the number of TRx that will be generated from these new patients and the refilled rates (from analogues) on a monthly basis in the first 24 months of launch.
   - TRxs = NBRxs + Refilled Prescriptions

3. **Revised the TRx Forecast**
   The Launch Excellence team adjusted and refined the client's TRx uptake curve to ensure that the monthly TRxs, monthly new patient acquisition and monthly physician initiation were reasonable and attainable. The revised forecast essentially became the client's new forecast.

Case Study 2

Using the NBRx Dynamic Simulator to solve for the number of new patients and physician initiators required, the Launch Excellence team discovered that the shape of the TRx uptake curve required fine tuning in order to achieve the monthly forecast.

Method

The Launch Excellence team performed the following analyses:

1. Starting with the client's monthly TRx forecast and refill rates from analogues, the Launch Team determined the number of new patients (NBRx required) on a monthly basis to hit the client's TRx forecast for the first 24 months.
   NBRxs = TRxs – Refilled Prescriptions

2. Then the Launch Team refined the NBRx curve to ensure that the monthly new patient acquisition was realistic and attainable in consultation with the client.

3. Using the adjusted month new patients or monthly NBRx curve, the Launch team remodeled both the monthly TRxs and monthly physician initiators required, using both the NBRx Model and the Physician Model included in the simulator. As discussed above, The NBRx model is driven by the following equation:
   TRxs = NBRxs + Refilled Prescriptions

   The Physician Model is driven by the following equation:
   NBRxs = Physician Initiators x Physician Value

Results

The clients were able to quantify the potential of their new brands and better position them for success. The Launch Excellence team provided the clients with:

1. A new TRx forecast with a realistic uptake curve
2. Clear monthly NBRx and physician targets
3. New performance benchmarks including refill rates and physician productivity
4. A dynamic simulator to generate new targets if changes are made to the TRx forecasts

Conclusion

Aligning new patient acquisition and prescriber initiation to the brands forecast creates a holistic view of planning and execution by connecting the following three work streams:

- Brand valuation (forecast)
- Go-to-Market structure (salesforce size and targets)
- Performance tracking

To achieve this, the NBRx Dynamic Simulator was developed to seamlessly bring together an NBRx Model (New Patient Acquisition Analysis) and a Physician Model (Physician Initiation Analysis) on a platform that is very user friendly and allows you to:

- Easily pressure test your brand forecast
- Run scenarios to determine the sensitivity and robustness of the underlying assumptions
- Develop key performance indicator for tracking a new launch or an in-market product
- Develop a new TRx forecast based on NBRxs

The methodologies and the assumptions driving the analyses are based on studying over 500 brands mainly in the 'Retail' space (products flowing through the traditional retail and mail order pharmacies). We believe that products in the 'Buy & Bill' space (products flowing through and administered by healthcare providers at their facilities) will also behave the same as the retail products we have analyzed.

We are in the process of doing the research with Buy & Bill products to confirm our belief. Therefore at the moment the tool should be used mainly for analyzing and brand planning around retail products. Future versions will be updated to also handle Buy & Bill products.

About the Author

***Ester Oben Etah**, Principal, Launch Excellence, IQVIA, has over 20 years of experience working in the pharmaceutical and biotechnology space both in the US and internationally. With her medical and scientific background, she brings therapeutic area expertise as well as an in-depth understanding of the rapidly changing dynamics of the industry.*

While at IQVIA, Ester developed experience in market assessment & forecasting and patient level data analysis. In addition, Ester has experience in product portfolio valuation & decision analysis, product licensing & deal valuation, mergers & acquisition analysis and life cycle management.

The Power of Feature Engineering Automation

Jean-Patrick Tsang, PhD & MBA (INSEAD), President of Bayser

Abstract: Feature engineering is the key to predictive modeling. If predictive performance is what you seek, endow your predictive model with powerful features. The problem is feature engineering is a manual process and essentially a hit-or-miss proposition. We developed a system that automatically discovers relationships in the data. Our system discovers the Newton's universal law of gravitation, the Gas equations, Kepler's 3 laws of planetary motion, Kleiber's law of metabolic rate, and the like. By offering a constant supply of powerful features for the human modeler to choose from, our system speeds up and enhances the feature engineering process, resulting in a significant boost in the accuracy of the predictive model.

Keywords: Feature automation, Genetic algorithm, Automatic relationship discovery, Panorama, Man-machine collaboration

1. Introduction

Whatever the accuracy of the predictive model, it is almost always not high enough from the moment it sees the light of day. That's the bane of the predictive modeler and it will keep the predictive modeler fussing over for a while. There are three places the modeler will look into to boost the accuracy of the model. One is to acquire new data sources to increase the number of variables in the model. Two is the algorithm, and this means either tweaking the hyper parameters of the algorithm or trying out new algorithms. Three is feature engineering, and it is the place that holds real promise once new data assets have been acquired and the final algorithm has been chosen and tuned.

Feature engineering is about creating new combinations of variables to explain the outcome variable. It is a very manual process and essentially a hit or miss proposition. Needless to say the deeper the understanding of the problem, the more likely the modeler will come up with clever features for the job. But there is no guarantee of success. The modeler may toil for a while and not have a eureka moment. If the dry spell lasts long enough, the modeler may quit in frustration, thereby sealing the fate of the predictive model.

Pause for a second and reflect on how feature engineering is carried out. Indeed, there is something very wrong with this picture and it is the reliance on inspiration, which we all know, cannot be brought about on demand. The overall task is to automate predictions without manual intervention and yet our approach is so manual. It's as if we had forgotten that we could use automation to help with feature engineering.

In this paper, we propose an approach that redresses the situation by automating the very process of generating features. To that end, we employ an evolutionary algorithm with a couple of adaptations to leverage the specific nature of predictive modeling. We tested our approach on different types of problems to ensure it can find interesting features. In the healthcare area, our algorithm discovers that Weight divided by Height squared, which is the BMI, is a great way to predict the weight profile of a person. In the world of physics, it discovers Kepler's 3 laws of planetary motion, Newton's law of universal gravitation, and the like. In biology, it discovers Kleiber's law which states that an animal is metabolically more efficient as it grows bigger.

Figure 1: Performance With and Without BMI

Algorithm	Scenario I (w/o BMI)	Scenario II (w/BMI)
Boosted Trees	0.92	1
Random Forests	0.90	0.99
Linear SVM	0.90	0.99

Our recommendation is to deploy our approach not as a replacement of the experts involved in the development of the predictive model but in conjunction with them. This is for two reasons, none of which being political correctness. The first one is to leverage the expertise of our experts more effectively and the second one is to guard against the possibility that the features our algorithm comes up with are based on noise rather than signal.

2. Problem

Why is it that the accuracy of a predictive model is less than perfect? It's because there are variables that are missing in the model. Actually, the correct term here is features. Indeed, features are variables or combinations of variables that the model uses to predict the outcome variable. Now, all the right variables may be present in the system but if the winning combination of variables is not made explicit, the model will have trouble predicting the outcome variable with accuracy.

Let us illustrate this point with a simple experiment. First, we provided the height and weight of several people and asked the model to predict who is underweight, normal, or overweight. We used Boosted Trees, SVM, and Random Forests, and they all did okay. We then repeated the experiment and this time we explicitly included weight divided by height squared. Needless to say all the models immediately registered perfect accuracy. See Figure 1. The point here is that unless the model is given the feature explicitly, it will not on its own find the winning combination of variables. This is what feature engineering is all about. By the way, the gist of our approach is to generate these features automatically.

In this BMI example, the winning feature is obvious to anyone that has a modicum of knowledge in health and fitness. But what if that's not the case? Let's consider an example from the Kaggle competitions where the winning feature is less obvious. The problem is to identify the type of car that has the highest resale value. In other words, what feature of the car does it? Is it the brand, the leather seats, the stereo system, the size of the wheels, the comfort of the ride, the size of the trunk? Or something else? The answer of the winning team, it turns out, is the color of the car but with a twist. Indeed, the color has to be unusual for that type of car, for instance, bright orange for a mid-size sedan. The reason actually has nothing to do with the color per se. It's just that people that really love cars buy cars of unusual colors and since they take very good care of their cars, the car retains its value and commands a high price at resale.

We have all tapped into our knowledge of the subject matter to come up with clever features. In one project where we had to predict which physicians are most likely to prescribe a drug, we used the fact that the drug is expensive and developed a drug price sensitivity index of the physician that looks at the 80th percentile of the price of the drugs the physician prescribes. This turned out to be an excellent predictor. In another project, we had to predict which physicians to target. We used data from the Sunshine Act to develop an allegiance index

that in essence indicates if the physician takes money from only one pharma company or from a plethora of companies. The allegiance index turned out to be an excellent predictor. Warren Buffet, even if he is not a student of predictive modeling, came up with a very interesting feature to predict the gyrations of the stock market. It is the ratio of the market capitalization to GDP (Gross Domestic Product) and is known as the Buffet indicator in investment circles.

All the features we discussed so far are the result of some insight or hunch. What if the insight is misguided or the hunch unlucky? What if after several fruitless attempts, the modeler and companion experts run out of ideas? Well, when that's the case, the whole process grinds to a halt and that's the end of the predictive model as far as its accuracy is concerned.

What the system is cruelly missing is a mechanism that steps in and proposes a stream of features for the expert to consider especially during the dry spells. The features do not have to be perfect. Even when the feature is off, it may serve as a springboard to get the creative juices of the expert to flow again. Indeed, by submitting something concrete to the expert to react to, we put the expert in a mental state that makes it easier to think of tweaks or explore ideas that the tentative features triggered.

In addition to the paradoxical situation we are in, which is to use a manual process to solve the key component of an automation problem, there is yet another problem. We do not know when to stop looking for potentially better features. On the one hand, we need to keep on looking as the feature may be right under our nose and we simply have not spotted it yet. On the other hand, there is no promising feature to uncover. Any further exploration is a pure waste of time.

What we need is a mechanism that provides feedback as to where we are in the exploration process. Should we keep on looking for clever features, or stop? This mechanism should keep tabs on our progress and, when appropriate, step in to indicate that there are no obvious stones that have been left unturned. The likelihood of discovering a new feature is so slim that it is best to discontinue the search.

3. Our Approach

Our approach operates under the following assumptions. First off, it recognizes that the variables that the predictive model works with come from different databases and that each database has a different layout (e.g., Sunshine Act, Physician Prescriptions, Referrals, Patient Adherence). Second, the performance of the ML algorithm (e.g., Logistic Regression, Boosted Trees, SVM, Random Forests, etc.) is valuable feedback that can be harnessed to guide the search for interesting features. Not only does the ML algorithm indicate the performance of the features as a group, it also provides the performance of individual features.

Our system automates the generation of features through two modules: the Extractor and the Mixer as shown in Figure 2.

One copy of the Extractor is assigned to each database and its role is to construct features by running algorithms on the database. Which algorithm is run depends directly on the database of interest. If it's a Referrals database, features that are built include the degree of the physician (number of physicians the physician is connected to), the PageRank score of the physician (this centrality measure captures not only the number of connections of the physician but also the importance of the physicians the physician is connected to), the closeness of the physician (average number of connections to reach any physician in the entire network), and

Figure 2: Architecture Including the Extractor and Mixer

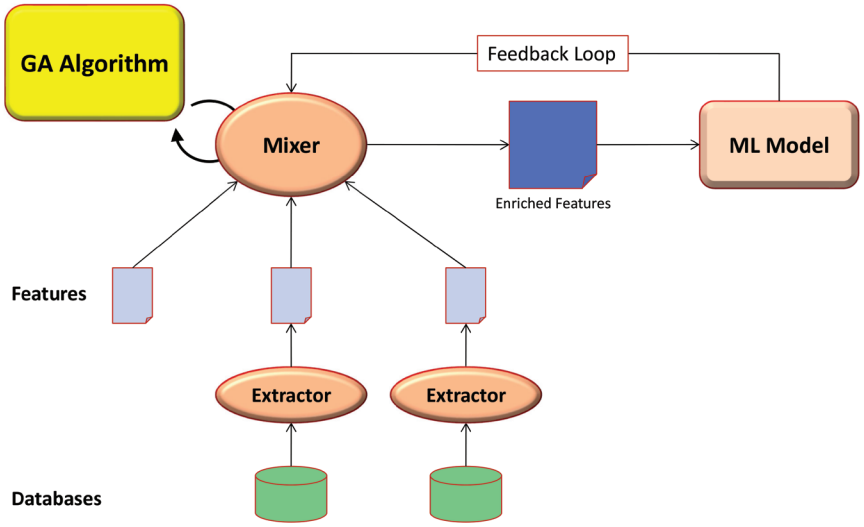
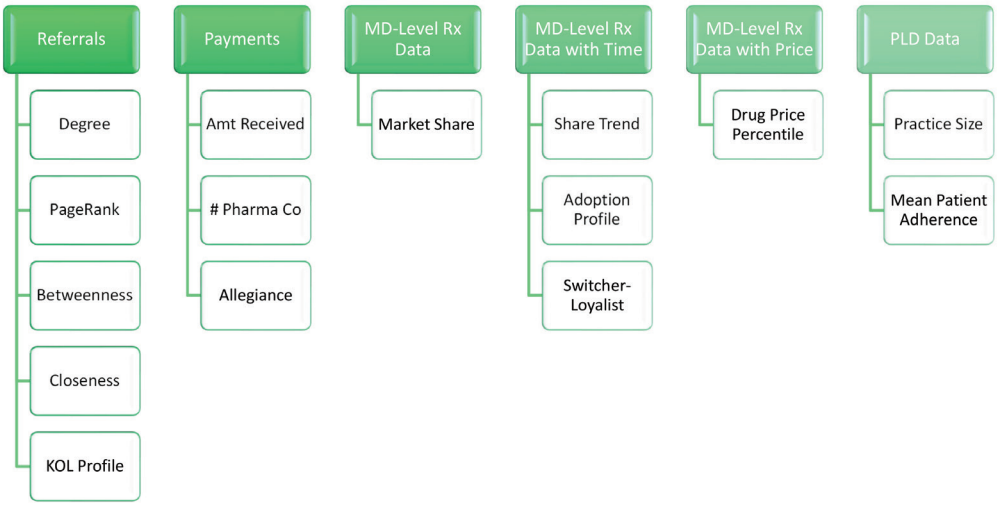


Figure 3: Examples of Operators Used by the Extractor



the like. If it's the Sunshine Act Open Payments database, the features include a drug allegiance index that in essence indicates the number of pharma companies the physician receives money from in regards to a given drug. The same goes for all the other databases. See Figure 3.

Once the Extractors are done building features from their respective databases, the features are fed into the Mixer. The Mixer then picks out features from the long list of candidate features and uses a bag of Operators (see Figure 4) along with the crossover, mutation, and mirror image procedures to construct new features. These features are then fed to the ML model. The Mixer then receives feedback from the ML model regarding the performance of the features it just provided. It then uses the feedback to construct the next generation of features. This process is repeated over and over again until we barely see an improvement in the explanatory power of the features. At this point,

Figure 4: Examples of Operators Used by the Mixer

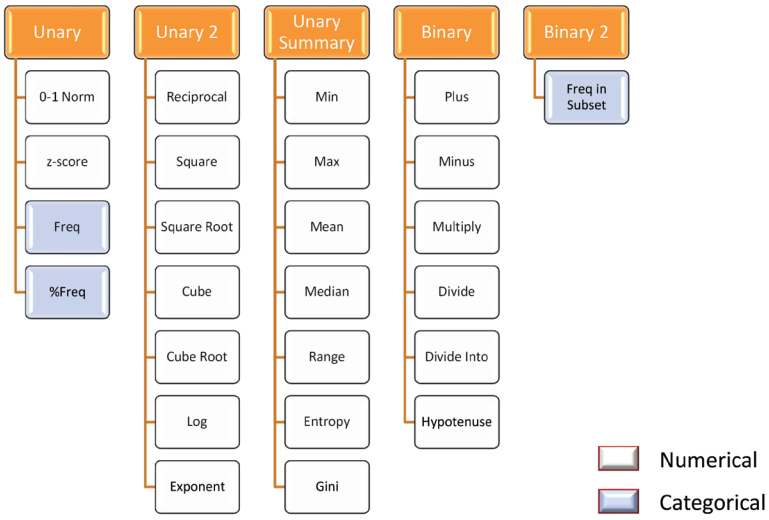
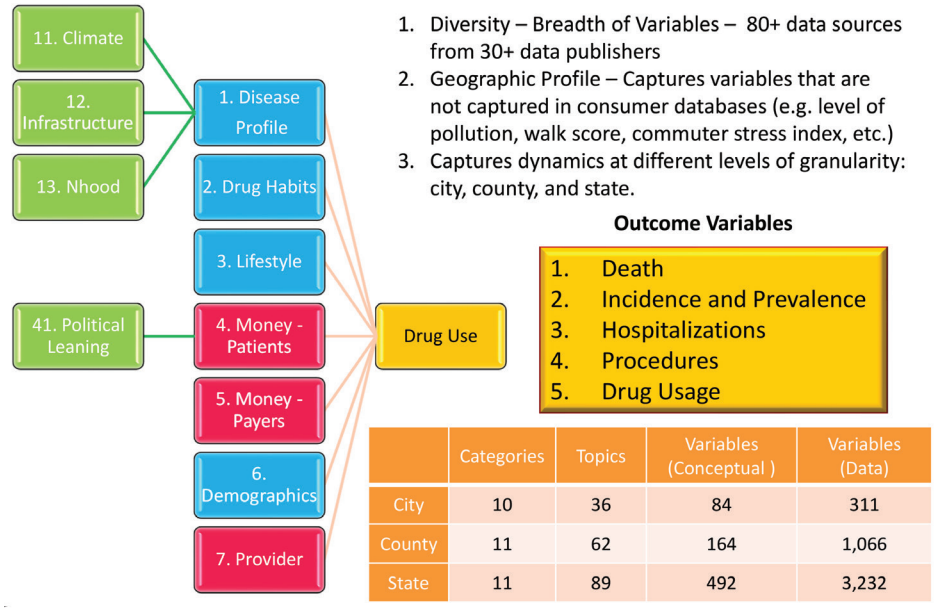


Figure 5: Overview of Panorama, the Companion Data Asset of Predictive Models



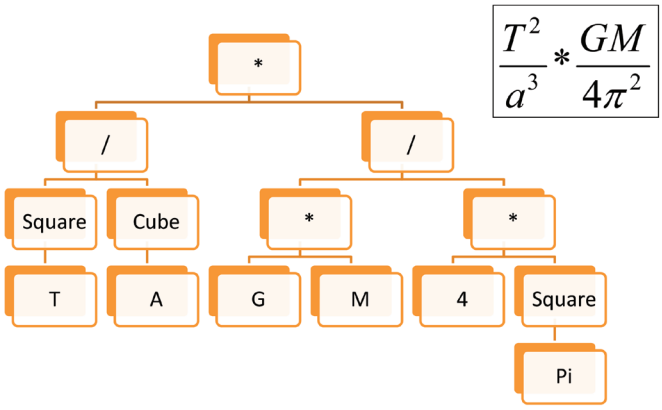
	Categories	Topics	Variables (Conceptual)	Variables (Data)
City	10	36	84	311
County	11	62	164	1,066
State	11	89	492	3,232

we typically have a couple of interesting features on our hands. Nonetheless, we may be missing some very interesting features. To that end, we'll start the whole process anew several times to ensure that the search is conducted from different initial conditions. At this point, it is most unlikely that we have missed some very interesting features. In other words, this process is quasi complete for all practical purposes.

There is one requirement for our approach to work and it is that the variables should be large in number and broad in diversity. That's because novel features cannot emerge through combinations drawn from the same small pool.

To that end, we developed a companion database, Panorama, that we leverage in our predictive modeling projects. See Figure 5.

Figure 6: Tree Representation of an Expression



Panorama pulls data from several data assets, each one focusing on what we call a category: disease profile, climate, infrastructure, neighborhood, drug habits, lifestyle, money from the standpoint of the consumer, political leaning, money from the standpoint of the insurance, demographics, provider, and healthcare resource utilization (drugs, procedures, surgeries, and hospitalizations). Each category is made up of what we call topics. For example, the disease profile category comprises five topics: incidence/prevalence, mortality, oral hygiene, diagnostic testing, and mental health. Topics are made up of variables. Examples of variables include particulate matter pollution, walk score, commuter stress index, allegiance to pharmaceutical companies, indifference to drug pricing, heart rate at rest, temperature spread, and exposure to UV, to mention just a few. Information is captured at four geographic levels: State, City, County, and Zip. Currently, Panorama counts over 3,200 variables culled from 80+ data sources and 30+ data publishers.

An early implementation of our approach revealed two problems. First, the Mixer was coming up with very complex features at the detriment of simpler ones. Second, some of the features used overly complex expressions. Example: square of the square root of x. In other instances, the expression was a constant in disguise. For instance, $(x^2 - y^2)$ divided by $(x - y)(x + y)$, which is 1.

For the first problem, we used a regularization technique in keeping with Occam's razor. We assign a complexity score to each feature and have the generative process penalize features with high complexity scores. In our implementation, a feature is represented as a tree (Figure 6) and the complexity score is simply the number of nodes in the tree. For the second problem, we articulated rules that forbid the generative process from building unnecessarily complex expressions. The rules are not foolproof, as some of the expressions may be quite tricky to call out. Nonetheless, the deployment of these rules greatly cuts back on the creation of irrelevant features.

Another issue we faced is that a whole slew of features were missing. They were features that involve constants. A feature such as square root of $x + 2$ is never generated as the Mixer only combines features with other features and 2 is not a feature. The solution came to us once we realized that the generative process does not need to consider all constants, but only a handful such as 2, 3, π, and e. By conferring the

Figure 7: Features Discovered by the Automatic Feature Generator

Cnt	Relationship	Formula
1	Definition of BMI	$BMI = \frac{W}{H^2}$
2	Newton's Second Law of Motion	$F = ma$
3	Energy of a Photon	$E = h\upsilon$
4	Simple Harmonic Time Period of Spring	$T = 2\pi\sqrt{\frac{m}{k}}$
5	Simple Harmonic Time Period of Pendulum	$T = 2\pi\sqrt{\frac{l}{g}}$
6	Kepler's First Law - Ellipse	$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$
7	Kepler's Second Law – Equal Area in Equal Time	$\vec{L} = \vec{r} \times \vec{a}$
8	Kepler's Third Law – Time Squared and Semi Major Axis Cubed	$T^2 = \frac{4\pi^2}{GM} a^3$
9	Newton's Law of Universal Gravitation	$F = \frac{GMm}{r^2}$
10	Gas Laws – Boyle, Charles, Gay-Lussac, and Avogadro	$PV = nRT$
11	Euler's Polyhedron Formula	$V - E + F = 2$
12	Kleiber's Law of Metabolic Rate	$E = M^{\frac{3}{4}}$

status of variables to these constants, the generative process treats them as such. This approach worked well and the next batch of features now had constants.

Regarding the choice of the evolutionary algorithm, we went with GA (Genetic Algorithm) for two reasons. First, we know this algorithm well and have implemented several variations of it to solve client problems. Second, GA is better suited for discrete problems like the one we are dealing with here than newer and greatly successful algorithms such as PSO (Particle Swarm Optimization) and DE (Differential Evolution).

4. Results

We ran our algorithm on 12 data sets (Figure 7). For each data set, the predictor variables are the ones that appear in the equation. In the case of BMI, the variables are Height, Weight and BMI. As for the outcome variable, it is 1 for all the records, in other words, the vector 1. The task of the algorithm is to explain vector 1 using the predictor variables. We use RMSE (Root Mean Square Error) to gauge how far an expression is away from 1. In a noiseless environment, the target RMSE is 0. When the RMSE is very close to 0, this means that the algorithm has uncovered a relationship between the variables.

Some expressions contain not only variables but also constants. We handle this by giving the status of variables to these constants in the algorithm. That's the case of G in Newton's law of universal gravitation. That's also the case of 2 and π, which appear in the simple harmonic motion formulas and Kepler's third law for instance.

For each run, the size of the dataset is between 100 to 1000 data points. The algorithm discovers all the relationships in less than 10 generations. Each experiment takes less than 1 minute on a regular laptop.

What these experiments demonstrate is that if there is a relationship between the variables of a data set, our algorithm will find it. Now, predictive problems we encounter in the pharma world are not that cut and dried and the

relationships between variables not that crisp. This has two consequences. First, our automatic feature generator will most likely stumble on these relationships but may not recognize them as compelling, as the model does not register a significant drop in RMSE. Second, our algorithm may uncover a slew of relationships of lesser significance that crowd out the truly relevant ones.

Indeed, the ability of our algorithm to identify compelling features will weaken as the relationships between the variables in the data set become less obvious. This means we need to bring in a human being into the equation to see through the clutter and spot relevant relationships.

5. Conclusion

We saw that the current approach to feature engineering, the lifeblood of predictive models, dooms many projects to failure because of its over-reliance on inspiration, something that cannot be delivered on demand.

Our solution to the problem is a system that automates the feature generation process. We demonstrated through a dozen of experiments that if there is a clear relationship between variables of the data set, our algorithm will find it.

The problem though is that relationships that need to be uncovered in the pharma world are subtle and not easy to pin down. As such, they may go through the algorithm unnoticed and fail to make the cut as features of note.

Our recommendation is not to replace the human by the automatic feature generator, but rather to have them work as a team. Indeed, the human can see through the clutter and pick out interesting features that our algorithm may miss. The automatic feature generator may in turn spark new ideas in the human by delivering a constant flow of candidate features for the human to sift through and reflect on. Without this ongoing nudging, the human would probably still be waiting for that elusive eureka moment to break the dry spell.

Having the automatic feature generator and the human work side by side is undoubtedly a win-win.

About the Author

Jean-Patrick Tsang *is the Founder and President of Bayser, a Chicago-based consulting firm dedicated to pharmaceuticals sales and marketing. JP is an expert in patient-level data and related analyses ranging from longitudinal analyses to hospital-retail spillover, KOL identification, influence mapping, referral networks, molecular targeting, promotion response, and the like. JP publishes and talks on a regular basis and runs one-day tutorials several times a year. In a previous life, JP deployed Artificial Intelligence to automate the design of payloads for satellites, methaners, ethaners, and cruise-liners. JP earned a Ph.D. in Artificial Intelligence from Grenoble University and an MBA from INSEAD in France. He was four times an examiner for PhD dissertations. He was also the recipient of the 2015 PMSA Lifetime Achievement Award.*

References

1 Zheng, A. and Casari, A. Feature Engineering for Machine Learning - Principles and Techniques for Data Scientists, O'Reilly, ISBN 978-1-491-95324-2., 2018, 217 pp.

2 Bodenhofer, U. Genetic Algorithms: Theory and Applications, Lecture Notes, Third Edition - Winter 2003/2004, Fuzzy Logic Laboratorium Linz-Hagenberg, 126 pp.

3 Kachitvichyanukul, V. Comparison of Three Evolutionary Algorithms: GA, PSO, and DE. *Industrial Engineering & Management Systems*, Vol 11, No 3, September 2012, p. 215-223.

4 Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission, KDD'15, August 10-13, 2015, Sydney, NSW, Australia. http://dx.doi.org/10.1145/2783258.2788613.

5 Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E.B., Turaga, D. Learning Feature Engineering for Classification, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2529.

6 Katz, G., Chul, E., Shin, R., Song, D. - ExploreKit: Automatic Feature Generation and Selection, 2016 IEEE 16th International Conference on Data Mining (ICDM), 12-15 Dec. 2016.

7 Tsang, J.P. Deploying Machine Learning for Commercial Analytics, *Journal of the Pharmaceutical Management Science Association*, Spring 2018, p 79-91.

Advanced Targeting for TV and Digital Case Studies

Bob Doyle, Senior Director, Consumer, US Brand & Marketing CoE, IQVIA; Daniel Vun Kannon, Manager, Client Solutions, US Brand & Marketing CoE, IQVIA; Yong Cai, Director, Advanced Analytics, IQVIA

Abstract: Pharmaceutical marketers are facing increasing pressure to prove their marketing investments are driving business. To see success during campaign reporting, it is important that the campaign is set up to target highly qualified targeting segments to optimize media spend, eliminating waste, and therefore driving ROI. Thanks to advances in data management and analytics, pharmaceutical marketers can now leverage 1st, 2nd, and 3rd party data to develop these targeting segments and to report against. We will further discuss in detail how these different data sets were implemented into campaign setup for targeting and the results from that campaign (audience quality and ROI).

Keywords: Digital media, Television, Campaign optimization, Media measurement, DTC marketing, Advanced targeting

Background

Data Sources

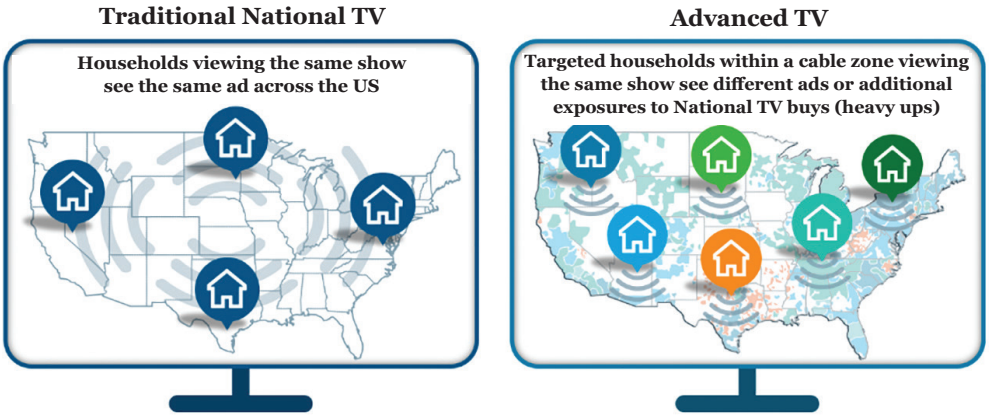
It is important that marketers leverage the most timely and accurate data sets when setting up targeting and measurement for a campaign to ensure the brand is reaching the most qualified audience available. Brands want to communicate their messaging to consumers who are most likely to switch to their brand and don't want to waste advertising impressions on consumers who would never switch or aren't eligible. A campaign targeting a qualified audience will drive higher ROI than a campaign targeted to the masses. However, not all data sets are created equal; for example, many closed looped data sets available on the market only update their datasets a few times a year. As you could imagine static data sets like these could impact targeting and ROI reports if the Rx activity isn't being captured within the campaign & reporting time frame. For example, many DTC campaigns will measure total Rx lift with an eye on new and continuing users. If a data set is only updated twice a year, you may miss the opportunity to target consumers during the campaign as they weren't identified as a qualified target in time. Furthermore, the campaign may have reached the right audience, but they weren't counted as a new or continuing patient as the update to the data set didn't coincide with the study period.

In addition to timeliness, it's also imperative to ensure the accuracy of the patient activity in the data as well. Many data sets will model the target consumer from small samples of patients and find commonalities amongst them. Those common traits are then bundled together to create a modeled consumer and from there targeting, measurement, test and control groups, etc. will all be based on this modeled consumer.

Modeled data sets can be useful for marketers when they need to reach a certain population size for measurement. Many campaigns focusing on rare diseases or conditions with low national prevalence usually find a need for these as the population samples are inherently low.

Since there are assumptions being made with modeled data, actual patient level data which provides insights into unique patients will yield more accurate audience targeting and

Figure 1: Traditional National TV vs. Advanced TV



measurement, as the test and control groups will be comprised of actual patients. Furthermore, some patient level data sets will provide a longitudinal view into the patient's medical history, allowing marketers to understand the entire medical history and burden that the patient is facing. For example, a pharmaceutical brand which treats diabetes may want to target patients with the corresponding diagnosis. However, no two diabetes patients are alike, and one patient may be on multiple drugs which may have adverse effects when combined with the Pharma brand's drug, whereas another may have a secondary condition such as a digestive issue which may also be treated by the drug, which would make them a high valued target for the brand. This is just one example of how accurate data can help refine audience quality by ensuring a brand doesn't waste targeted media on someone who would never be eligible anyway. Another way in which you can ensure the data source that you are using for targeting isn't hindering your marketing efforts is to ensure the population set is deep enough to avoid holdouts. No client will want to limit their marketing message against a qualified target for the sake of measurement. Instead, it's preferable to ensure your data is timely and accurate to avoid these situations.

Here we will illustrate how advanced targeting powered by patient level data helped brands refine their targeting and drive ROI.

Television

When assigning media budgets, pharmaceutical marketers continue to invest the lion's share of the media budget into television due to its high reach and prominent viewing area for media consumption. Traditionally television ads were served nationally to the whole USA or to smaller market segments known as DMAs (Designated Marketing areas). There are 210 DMAs across the USA which represent the various TV markets. Placing TV ads nationally or at the local DMA level worked well for advertisers for a while. However, as pressure began mounting to measure and optimize those investments, digital technology was implemented into the TV ecosystem to better target and measure TV. To answer this need cable companies developed the addressable TV solution. Cable companies began leveraging their subscriber's data and information to allow advertisers to deliver ads to them at the household level or to a whole cable zone which is a designated group of zip codes within a DMA where the cable company has a footprint (Figure 1). This technology isn't brand new but hasn't been fully leveraged by advertisers yet because the data set was still limited to the cable company's subscriber data which only told them the name, age, and address of the subscriber. Now data onboarding and management companies, such

as Liveramp and Adobe, enable the marrying of different data sets which allow marketers to combine their own CRM data with the cable subscriber data completing consumer profile enabling advanced TV targeting. Leveraging various data sources available for targeting and measurement is the primary focus of many marketers across all industries including CPG, automotive, pharmaceuticals etc. as the data sources allow refined targeting opportunities.

Digital

Investment into digital media is increasing rapidly due to the inherent ability of the medium to target, optimize, and measure consumers. Digital covers a very wide variety of mediums such as display, online video, search, video-on-demand, streaming audio and can even encompass digital placements in doctors' offices. Digital media is easier to track vs. analog mediums such as TV, radio, or print as consumers are tracked constantly when online whether through cookies, a piece of code that tracks that consumers browsing history, or through digital ID's provided by the user's device, i.e. smartphone, tablet, PC, etc.

Traditionally digital ads were placed directly onto sites where the target audience was most likely to consume media. For pharmaceutical DTC campaigns, this usually entailed targeting health endemic websites to reach qualified consumers as they were trusted sources of information with high audience composition of the target consumer. Ad effectiveness was generally rated on how many impressions were served (volume) and if the consumer clicked on the ad (engagement). However, these metrics are only proxies for actual real-world behavior and can't be used to determine whether a campaign drove lifts in awareness or incremental prescriptions (new or continuing).

Now, with contemporary data management technology and services, marketers can link campaign exposures to real world actions to determine lift and ROI. Furthermore, advances in ad-measurement & ad-fraud technology enable marketers to draw correlations between media performance KPIs and business outcomes, enabling them to optimize to higher performing partners and creatives within a campaign. Again, these advances wouldn't be possible without the data providing the foundation for these capabilities.

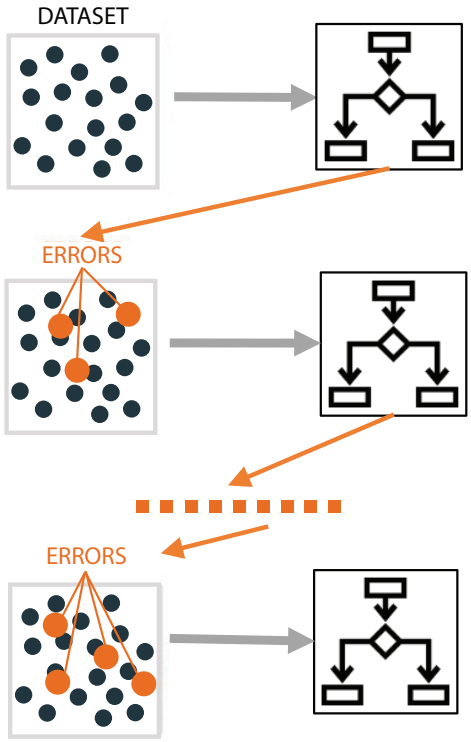
Methodology

Following is an approach to develop advanced targeting segments to be leveraged for digital and advanced TV campaigns when targeting at the household level.

Let's define US population as Ω, consumers from publisher(s) database be Ψ and patient population be Φ, where $(\Psi, \Phi) \subset \Omega$. Φ represents the 300 million patient set captured in either prescription (Rx) or claims (Dx) data. Let X be the crosswalk between the anonymous IDs of Ψ and Φ. Note that $(\Psi \cap \Phi) \subset X$ and X is a subset of Ω as well. Using market definition and disease related business rules, we can filter Φ into $\varphi 1, \varphi 2, \varphi 3, .. \varphi n$. φk represents the set of patients of interest with certain conditions (Dx) or potential users of a product (Rx).

Using the above definition, we build our advanced target list from 3 components: the high value, lookalike and low value lists. The high value target list is defined as $(\varphi k \cap X \cap \Psi) \cup \Delta$. Δ is the set of non-relevant patients. The non-relevant patients are included for privacy compliance to prevent any identification of a consumer based on the targeting they receive. For sensitive markets (e.g. mental health, sexual conditions, etc.), the high value/non-relevant ratio is set to at least 50%; and for non-sensitive markets, the high value/non-relevant ratio at least 20%. The low value list is comprised of patients who are excluded from

Figure 2: Extreme Gradient Boosting (XGBoost)



targeting, Ψ∩(Φ - φk). This is only used for suppression purposes in promotion activation. The lookalike patients are the predicted targets from the unknown patient set. Using φk and selected controls, we use consumer attributes and behavior variables such as clicks and topic keywords read in Ψ to predict potential targets.

The default model is Extreme Gradient Boost (XGBoost), which is a popular tool that often wins Kaggle competition. Our system is built upon Hadoop and Spark. The Pyspark pipeline enables us to have a very efficient predictive engine. The processing time can be reduced up to 3/4 compared to using a traditional Oracle system plus 3rd party statistical software. See Figure 2.

To achieve maximum targeting efficiency, campaign managers start campaign activation from the high value list and gradually move down to the lookalike patient list that is ranked by probability prediction. In this process, they suppress activation for any consumers shown in the low value list.

All patients and consumers identifiable information are either encrypted or removed by 3rd party. We only handle anonymous patient/consumer IDs and features that are HIPAA compliant.

Advanced Digital Targeting

We were tasked to help improve targeting on a single media partner within a client's campaign to prove that advanced targeting will help improve ROI. The brand wanted to focus on patients with a prior Rx within the category so we helped define that target by leveraging its patient level data per the methodology previously discussed. From there, the advanced targeting segments were cross-walked with a 3rd party consumer partner and then deployed for activation directly to the media partner.

Figure 3: Audience Quality

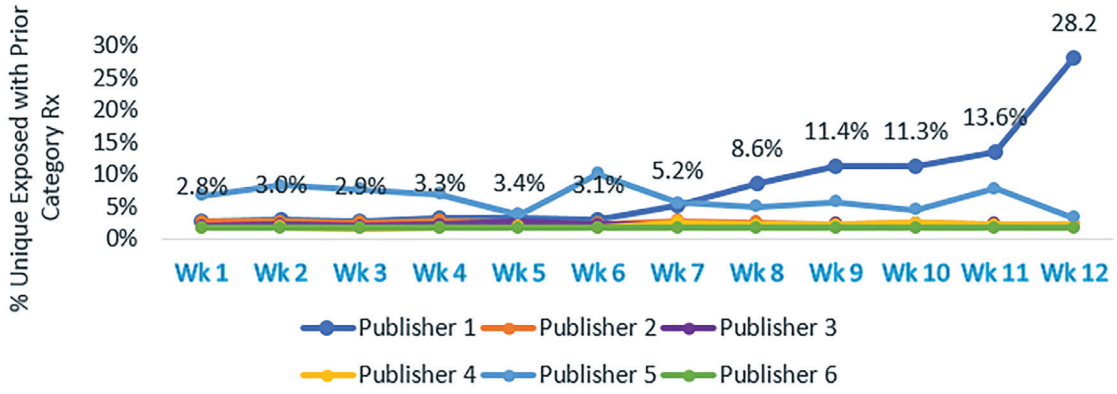
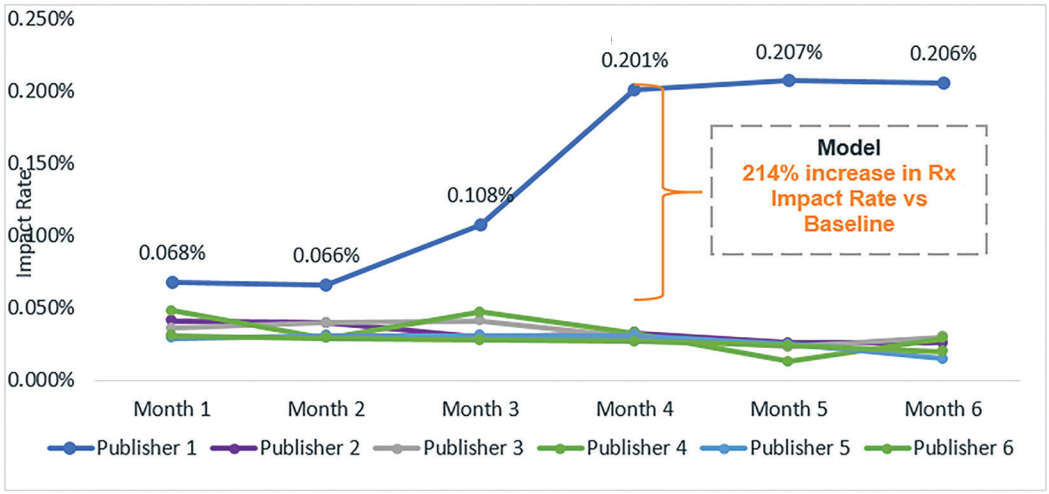


Figure 4: Impact Rate



In Figure 3, you can see the improvement to audience quality when the advanced targeting segments were deployed to Publisher 1 during week 6. Audience quality is defined as the % of the audience exposed who had a prescription for a drug within the defined category within the past 12 months. The audience is measured via a media tagging pixel and cross-walked via a privacy compliant manner to the Rx database.

By week 7, the campaign's ability to reach patients who had previously been treated with a product in the category of interest increased by 111%.

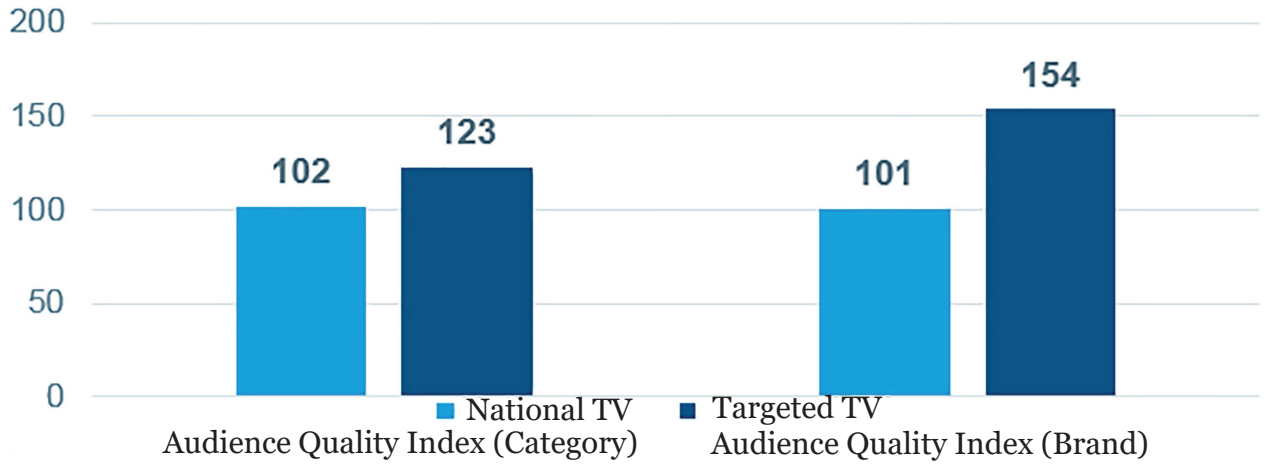
By week 11, the campaign's ability to reach patients who had previously been treated with a product in the category of interest increased by 404%.

At the end of the campaign, impact was measured to see if the high composition of qualified patients drove brand Rx lift. Test consumers were matched at a 1:1 ratio to control pairs using demographic, healthcare, and media behavior attributes to determine the matches. Impact Rate % is then defined as the % of exposed patients with an Rx fill or administration post ad exposure. In Figure 4, you will see that the publisher with the advanced targeting segments saw an increase in brand Rx lift of 214% vs. the other publishers in the campaign.

Advanced TV

We were approached by a pharmaceutical brand to help them understand if Advanced TV targeted to a cable zone or even a DMA level

Figure 5: Targeted vs. National TV Indices



was more effective in driving additional Rx vs. National TV. In order to prove that focusing on actual patients and ensuring the messaging to those households instead of reaching the entire nation was more effective, we needed to identify the households of value. We then analyzed all the longitudinal patient data available to them and indexed each cable zone against the nation to identify high valued zones where composition of patients was high, had a doctor who was able to write the prescription, and were likely to use a brand within the category.

Figure 5 illustrates the difference in audience quality between Advanced TV vs. traditional National TV. Audience Quality scores are indexed against the general population and reflect the likelihood to have had a brand or category Rx within the last year. The methodology laid out in the previous section is the same used to develop the audience quality scores in Figure 5. Since media was being targeted to cable zones and not specific patient households, no non-relevant patients were needed when developing the audience segment.

TV media was then deployed for ~3 months and was measured post campaign to determine new patient starts and adherence. Our vast data sets consist of patients outside of the cable zones featuring media, so no holdouts were needed as enough patients were identified outside of the cable zone footprint for a statistical significant test/control analysis on Rx lift. Consumers were matched at a 1:1 ratio to control pairs using demographic, healthcare, and media behavior attributes. Overall advanced TV drove a 3:1 ROI for the brand outperforming National TV, with a 25% New-To-Brand Rx lift and +7% lift for continuing patients.

Conclusion

Advanced Targeting for TV and Digital reaches audiences whose behavior shows a high probability that they – or someone in their household – have a condition that could be treated by the treatment advertised and will convert to a doctor's visit asking for the treatment. This solution is reshaping the way the industry spends their advertising dollars and is helping pharma companies rethink their approach to television and digital advertising spend.

About the Authors

Bob Doyle *is a senior executive with over 30+ years' experience, skilled in business strategy, marketing, business development and consumer/patient insights. Drawing on a broad background that includes experience with consulting, healthcare, Pharma, and brand-leading CPG organizations, his accomplishments include building business strategies, driving P&L, and leading companies, clients and retailers in implementing strategies that improve profitability and shareholder return.*

Daniel Vun Kannon *has recently joined the IQVIA team after primarily working for various media agencies in New York City. He brings 6+ years of experience of delivering cross-channel media strategies and tactics to various clients across the CPG and QSR categories. He started his career focused primarily in multicultural planning and insights but began to focus more on data-driven strategies as the landscape evolved. His role at IQVIA now is focused on utilizing IQVIA's vast data sets to develop marketing insights for clients as well as developing innovative solutions for brands, agencies and their partners.*

Yong Cai *is a Director, Advanced Analytics, at IQVIA. He has more than 10 years of experience in data mining, developing new methodologies and designing innovative solutions to address complex business issues. His research fields are machine learning methods, predictive analytics and Bayesian modeling. He has expertise in data driven applications such as disease detection and patient outcome research, multi-channel marketing, pricing and market access and media innovation.*

Dynamic Asymptote Approach to Measure Interaction Among Promotion Channels

Shubham Lahoti, Associate Director, Axtria; Varun Jain, Associate Director, Axtria; and Adarsh Gautam, Project Lead, Axtria

Abstract: Over the years, pharma marketing has begun using more and more digital channels to complement the traditional marketing channels (for example, rep visits, journal advertising, and conferences). Numerous digital channels have sprung up, owing to technological advancements and changing doctor and patient habits. Digital platforms such as emails, digital ads, and paid search (PS) are available for pharma companies to target doctors, based on their preferences. These channels are deployed not with an aim to substitute traditional channels, but as part of a synergistic multi-channel marketing strategy to complement the traditional channels. Thus, there is an interaction effect between the digital channels and the traditional channels. Measuring this interaction effect has certain challenges. How does interaction effect change as the investment mix changes? How does one split interaction effect among the interacting channels? Do channels interact in a step-wise manner or in a continuous manner? This paper assesses how prior approaches measure interactions and their shortcomings. It then proposes dynamic asymptote approach to overcome those shortcomings. It will also evaluate various scenarios in which this approach leads to a better marketing strategy decision compared to prior approaches.

Keywords: Pharmaceutical, Marketing-mix analysis, Budget optimization, Interaction, Promotion channels, Dynamic asymptote

1. Introduction

Today's increasingly digital society is causing an evolution in pharma marketing strategies. Effective implementation of various digital marketing techniques along with the traditional methods has the potential to expand market reach, better engage doctors and thus provide a higher return on investment (ROI). However, effective utilization of health care professional (HCP) digital channels requires a precise tracking of their sales effect and ROI. Measuring ROI can be a tricky business because for most of the high value doctors (who are already exposed to rep calls), these channels are aimed at increasing brand awareness more than affecting the final prescription. They interact with traditional channels to affect sales. An accurate ROI assessment requires a clear measurement of this interaction effect. Analytics companies have deployed various methods to measure this interaction, but each method has its own limitations.

Thus, the goals of this paper are as follows:

- Given the expansion of additional promotional instruments, such as numerous digital channels: HCP digital ads, PS, social media (SM), and brand website and emails, emphasize the importance of measuring interaction among these channels and digital channels for marketing budget optimization.
- Assess prior approaches of measuring interactions and understand their shortcomings.
- Propose dynamic asymptote approach to measure interaction effects and understand how it is superior to prior approaches.

Figure 1: Channel Pathways Influence Doctors to Write the Final Prescription

Figure 1a

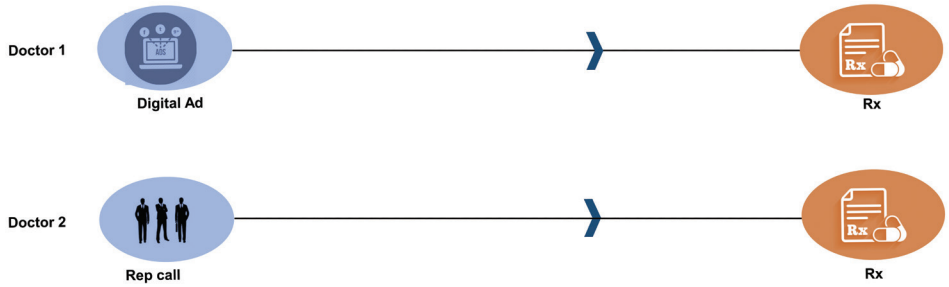
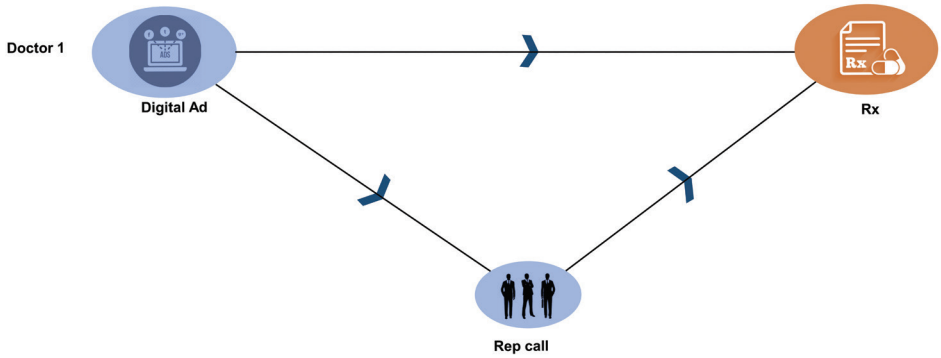


Figure 1b



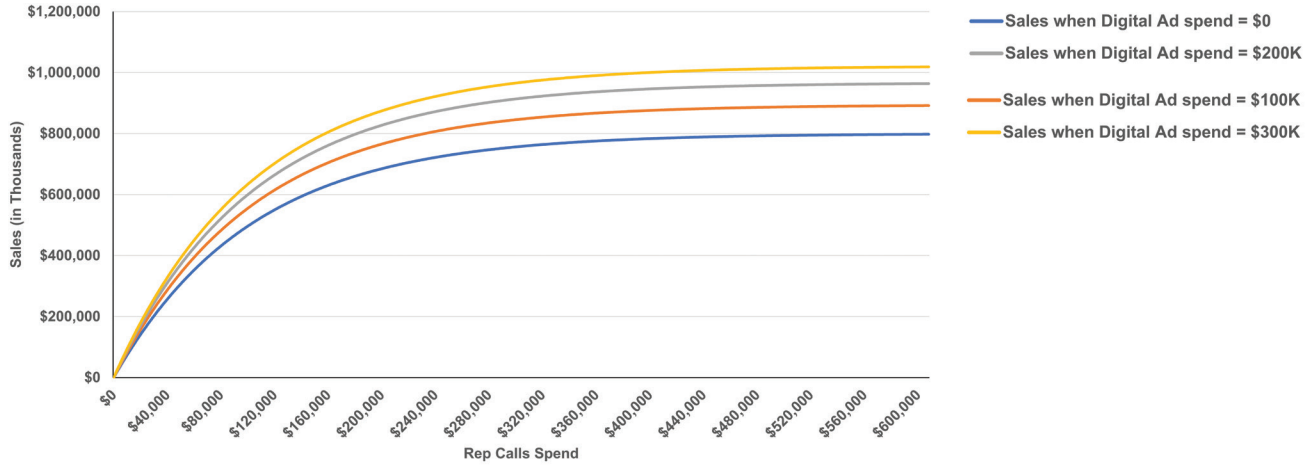
- Understand strategic benefits of the dynamic asymptote approach in making better budgeting decisions.
- This paper will finally propose areas of future enhancements to deal with the rapidly changing pharma marketing landscape.

The prescription happens after the doctor traverses through a complex 'pathway' of channels, where the doctor is becoming more and more convinced about the drug along the way. Figure 1 represents the journey of doctors through a typical array of channels – HCP digital ads, brand website, and rep calls before they write the final prescription. In Figure 1a, an HCP digital ad is exposed to doctor 1 and a rep call is given to doctor 2. In Figure 1b, the same doctor 1 is exposed to both the channels.

In Figure 1a, the overall effect is the sum of effect on doctor 1 and doctor 2. However, in Figure 1b, since the same doctor 1 is visited by a rep after seeing an HCP digital ad, they might remember the ad, resulting in them being more likely to prescribe after a rep visit. This means that the channels interact with one another to influence a doctor to prescribe a drug. In a sense, there are synergies such that the final delivered effect of two channels is greater than the sum of their individual effects.

Therefore, in most cases, the integrated marketing strategy not only increases customer engagement and expands market reach, but also provides a higher ROI.[3] It is important to measure interaction effect between marketing channels, so that their sales effect and ROI are accurately assessed. An accurate ROI will give all the channels their due credit and result in an optimal budget allocation.[4]

Figure 2: Rep Calls Step-wise Response Curves for Different Levels of Investment in HCP Digital Ads



1.1 Prior Approaches to Measure Interaction
Various approaches have been adopted in the past to measure interaction effect. However, there are challenges in incorporating them. Sub-sections 1.1.1 and 1.1.2 below lay out two approaches and the challenges associated with them.

1.1.1 Structured Equation Modeling[5]
Structured equation modeling (SEM) involves running two models sequentially – 1) primary model to calculate the effect of all channels assuming their effect is additive, and 2) pathway model to calculate the correlation between interacting channels.

[1] Primary model: $Y = \beta_1 + f(\text{Ch } 1) + g(\text{Ch } 2)$

[2] Secondary model: $\text{Ch } 1 = \beta_2 + A*\text{Ch } 2$

where,

Y = Dependent variable of the primary model

β_1 = Intercept of the primary model

β_2 = Intercept of the secondary model

f(Ch 1) = Transformation function for Ch 1

g(Ch 1) = Transformation function for Ch 2

A = Asymptote of Ch 2 in the secondary model

Using the secondary model, some of the effect of Ch 1 is credited to Ch 2, since Ch 2 is assumed to be driving Ch 1. The choice of the secondary model is dependent upon the modeler's assumption on which channels interact with one another and which one is "cause" and which one is "effect".

1.1.2 Step-wise Response Curve Modeling[6]
This approach assumes the channel interaction to be working in steps. For example, when the spending on HCP digital ads is increased, the response of rep calls would increase in a step-wise manner. Figure 2 represents step-wise response curves that capture interactions between HCP digital ads and rep calls.

This solves the problem of capturing interaction effects at various levels of investment mix and does not assume interaction to be a linear term.

1.1.3 Shortcomings of prior approaches (SEM and Step-wise response curve modeling)
Both these approaches have certain challenges. Some of the challenges of SEM are listed below.

1. *Ignores collinearity between interacting channels in the primary model and might result in overfitting***.** Therefore, it goes against the premise in the primary model that Ch 1 and Ch 2 are independent of each other.
2. *Assumes the relation between interacting channels to be linear, which is dependent on the starting point of the mix***.** In reality, the relation will change for different levels of investment mix. As the investment in one channel goes up, its effect on the other channel will not increase at the same rate.
3. *Presents an interpretability problem.* This approach interprets interaction as a one-way phenomenon, where effect of one channel is credited to the other. In reality, interaction is the overall synergy created when two channels work together. Hence its effect should be distributed across both the channels.

Also, the step-wise approach has a challenge that the number and size of the steps to be considered is a subjective one. In certain cases, observational data on campaigns clearly indicates that promotion increases in a step-wise manner and hence the effect also follows these steps. However, for most of the non-personal channels, promotional units are observed to be continuous, and hence should have a continuous effect. In such cases, interaction is not really a step-wise phenomenon. It is a continuous drive that pushes up the effect of a channel, as promotion in other interacting channels increases.

2. Dynamic Asymptote Approach Overcomes Limitations of Prior Approaches

Channels really interact in a continuous manner. As soon as promotion in one channel goes up, it creates a stir in the market. Doctors become more likely to prescribe, and other channels become more effective, even if to a minor extent (section 3.1 on the benefits of dynamic asymptote approach elaborates how measuring continuous interaction is better than measuring step-wise interaction). Going back to the previous example of HCP digital ads versus rep call combination, when there is a higher spend in HCP digital ad campaigns, rep calls will be more effective.

2.1 Modeling Construct

It is generally observed that rep calls have a diminishing response on the NRx (i.e., marginal response will keep on decreasing as the spending on the input variable is increased). There are several transformations that can be used to model this effect, for example negative exponential, logarithmic, and power transform. Assuming rep calls have a negative exponential response to NRx, effect of calls can be represented as the following expression:

$A_{Call}*[1-EXP\{-C_{Call}*X\}]$

where,

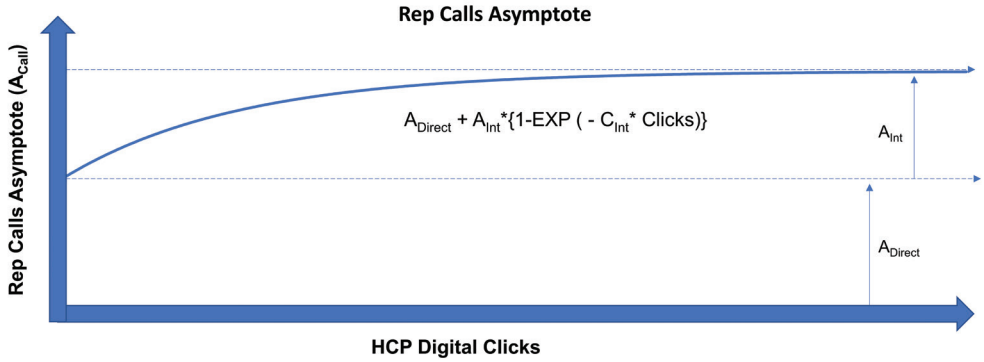
A_{Call} is the asymptote

C_{Call} is the curvature

X is the number of calls.

Continuous interaction means that the effect term ($A_{Call}*[1-EXP\{-C_{Call}*X\}]$) should be an increasing function of HCP digital ad spend. This could happen when either A_{Call} or C_{Call} increases with an increase in HCP digital ad spend. The maximum effect of calls goes up with increased spending on HCP digital ads. This can be achieved by defining A_{Call} in terms of HCP digital ad clicks. After establishing that A_{Call} should be a function of clicks, the nature of the function needs to be determined. Now, when clicks go up, either the same doctors are seeing more ads or newer doctors, who were not earlier targeted because they were low value, are seeing ads. In both the scenarios, the relation between A_{Call} and clicks must be

Figure 3: Rep Calls Asymptote Versus HCP Digital Clicks



that of diminishing return, with a saturation point. Such a relation is best explained using a negative exponential curve. Figure 3 illustrates this relationship using a sample negative exponential response curve.

Thus, A_{Call} can be represented as follows:

$A_{Direct} + A_{Int}$*{1-EXP(-C_{Int}*Clicks)}

where,

A_{Direct} = Asymptote when there is no interaction effect or when there are zero HCP digital clicks

A_{Int} = Asymptote of the interaction term

C_{Int} = Curvature of the interaction term

It can be observed that rep calls asymptote is increasing with an increase in number of clicks for the HCP digital ads. The minimum asymptote of rep calls will be observed at zero HCP digital clicks.

Thus, the total effect of calls will be specified as follows:

[3] *Calls Effect* = [{$A_{Direct} + A_{Int}$*{1-EXP (-C_{Int}*Clicks)}}* {1-EXP(-C_{Call}*X)}]

This effect is illustrated in the results (Figure 5 in section 2.3.1 below), after running the model on actual data.

Interaction is the overall synergy created when two channels work together. Hence, its effect could show up in term for calls effect (equation [3] above), or a similar term for HCP digital ads effect with an interaction component for calls.

[4] *HCP Digital Ad Clicks Effect* = [$A_{Clicks\_Direct} + A_{Int}$*{1-EXP(-$C_{Int}$*X)}* {1-EXP(-$C_{Clicks}$*Clicks)}]

It is theoretically possible that all channels interact with all other channels. Hence, the total number of interaction terms could go up to nXn, n being the number of channels. In order to decide how many interaction terms to use, a theoretical model needs to be designed before the empirical model. The theoretical model would explain all the causal relationships between the promotion channels. Interaction terms can explain these relationships. Once the theoretical model explains all such hypothetical relationships and possible interactions, an empirical model can be designed to check which of these relationships are statistically significant.

2.2 Benefits Over Prior Approaches

Dynamic asymptote approach overcomes the following challenges posed by prior approaches:

- It doesn't have the multi-collinearity challenge that SEM poses, as seen from the variance inflation factor (VIF) results (in section 2.3 below).

- It doesn't assume interaction to be a linear or a step-wise phenomenon. It measures interaction as a continuous effect, and hence enables measuring sales effect and ROI at a more granular level. This could lead the marketer to choose an investment mix that results in a better ROI (illustrated in an example in section 3.1 below).
- It is easier to interpret than prior approaches. Since interaction is measured as a separate term, it is simpler to allocate the effect to one of the two interacting channels. Effect split is explained in section 2.3.2 below.
- Since the model defines interaction term as a dynamic function, it works for any level of investment mix of interacting channels.
- This approach eliminates decision-subjectivity in determining the number of interaction steps and step-size.

2.3 Results and Summaries

A dummy cross-sectional, pooled time-series and panel dataset was created to mimic several actual client datasets to illustrate the dynamic asymptote approach. The dataset had promotion values at the physician and month levels. It was hypothesized that the dependent variable, NRx, will be dependent on promotion tactics as well as the lagged value of dependent variable. The following promotional data was available – number of rep calls, number of HCP digital ad clicks, webinar attendance, and number of emails delivered. A regression model was run with these variables, along with interaction terms, as explained in section 2 above. The interaction term was significant only for interaction between rep calls and HCP digital ad clicks. When the other interaction terms were introduced in the model, p-values for a few variables were observed to be above 10%. Empirical model design was specified as follows:

$$[5]\ NRx = \beta_0 + \beta_1 * Lagged\_NRx + \beta_2 * (1 - e^{-1 * curvature\_1 * Calls}) + \beta_3 * (1 - e^{-1 * curvature\_1 * Calls}) * (1 - e^{-1 * curvature\_3 * Clicks}) + \beta_4 * (1 - e^{-1 * curvature\_2 * Clicks}) + \beta_5 * Webinars + \beta_6 * Emails$$

where coefficient parameters represent effects on the following explanatory variables,

β_0 = Intercept

β_1 = Lagged NRx

β_2 = Asymptote of the calls

β_3 = Asymptote of the interaction term

β_4 = Asymptote of the HCP digital ad clicks

β_5 = Asymptote of the webinars

β_6 = Asymptote of the emails

Table 1 shows the estimates and standard errors for the independent variables, for the modeling construct that resulted in acceptable results.

In this case, it was observed that only rep calls and HCP digital programs interact with one another, with high significance, and a reasonable fit. The acceptability of the model fit was also validated by checking the residual plots. Figure 4 shows the plot of residuals versus the NRx term.

It is evident that the residuals are centered on zero and do not demonstrate any systematic pattern. Auto-correlation was also checked by regressing residuals versus lagged residuals. This regression did not give significant results, as seen in Table 2. This demonstrates that the model is not auto-correlated.

2.3.1 Dynamic Response Curves

Response curves for any channel (as described in section 3) will not be a 2-dimensional (D) curve.

Table 1: Outputs from Regression Model[7]

Dependent Variable: NRx				
Explanatory Variable	**Unstandardized Coefficients**	**Standard Error**	**Standardized Coefficients[x]**	**VIF**
Intercept	0.50***	0.06	-	-
Lagged NRx (-1)	0.66***	0.01	0.70	2.21
Asymptote of calls	0.34***	0.03	0.07	1.03
Asymptote of interaction term	0.07	0.04	0.01	1.01
Asymptote of digital ad clicks	0.01***	0.00	0.26	2.14
Asymptote of webinars	0.20*	0.10	0.01	1.03
Asymptote of emails	0.08***	0.01	0.07	1.16
*p≤0.05; **p≤0.05; ***p≤0.001				

Model Statistics:	
N	2,088
Multiple R-squared	0.9065
Adjusted R-squared	0.9062
Residual standard error	1.794
F-statistic (6, 2081)	3,363

[x] *Unstandardized coefficients were standardized by scaling the variances of all the dependent and independent variables to 1. Therefore, standardized coefficients indicate how many standard deviations a dependent variable will change, when the standard deviation in the predictor variable changes by 1.*

Figure 4: Plot of Residuals Versus the NRx Term

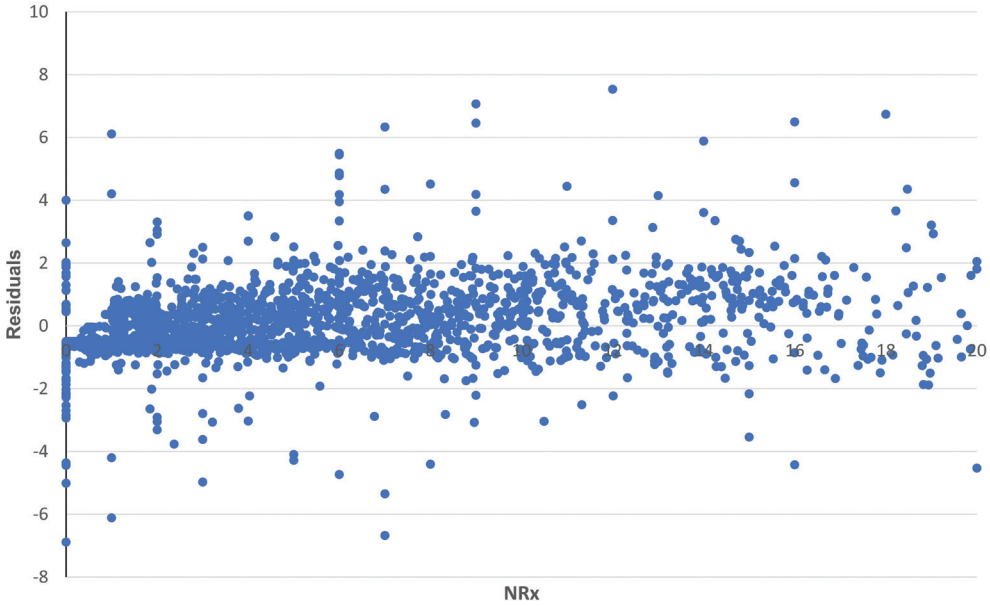
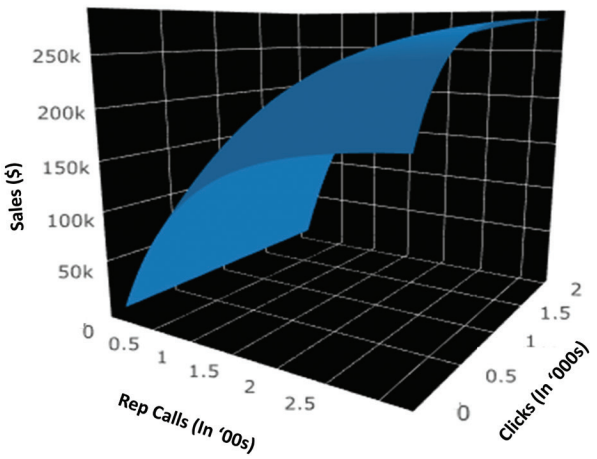


Table 2: Coefficients for the Regression of Residuals Against Lagged Residuals

Coefficients:		
	Estimates	**p-values**
Intercept	0.00	0.994
Residuals [-1]	0.02	0.305

Figure 5: Response of Calls on NRx Varying with Investment in HCP Digital Ads[8]



In the presence of multi-way interactions, the response would be n-dimensional. Figure 5 represents a 3-D chart (i.e., a response curve of a channel with two-way interaction). X and Z axes represent spends in two interacting channels (Rep Calls and HCP Digital Ads) and Y axis represents the sales that will be generated by these promotions.

It is observed, that the response to rep calls increase as the investment in HCP digital ads goes up. The rep call axis and effect axis at zero depth represents the response curve when the spending on HCP digital ads is zero. As the spending on HCP digital ads is increased, a third dimension is added to the 2-D response curve and the asymptote of the response curve increases. In this example, the increase in asymptote is driven by HCP digital ads. It is apparent that 2K HCP digital ad clicks will take the rep calls response to the maximum (asymptote). The absolute return will be a point on the 3-D Cartesian-Coordinate-System that will be driven by spending on both HCP digital ads and rep calls.

2.3.2 Direct and Indirect Effect of Interacting Channels

Dynamic asymptote approach enables splitting the effect of any channel into direct and indirect effect, since asymptote comprises fixed component and a component dependent on another channel. The effect term,

$A_{Call} + A_{Int}*\{1-EXP(-C_{Int}*Clicks)\}* \{1-EXP(-C_{Call}*X)\}$

can be split into two parts. Since the effect component,

$A_{Int}*\{1-EXP(-C_{Int}*Clicks)\}* \{1-EXP(-C_{Call}*X)\}$

is driven by clicks, it should be attributed to HCP digital ads, even though it shows up through calls. Combined effect of HCP digital ads is the sum of this direct and indirect ads.

Figure 6. Bifurcation of Total HCP Digital Ad Effect

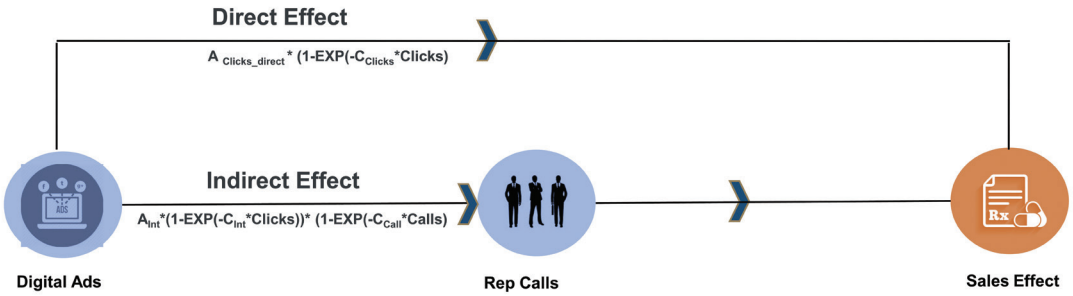


Figure 7. Promotion Effect Attribution to Channels

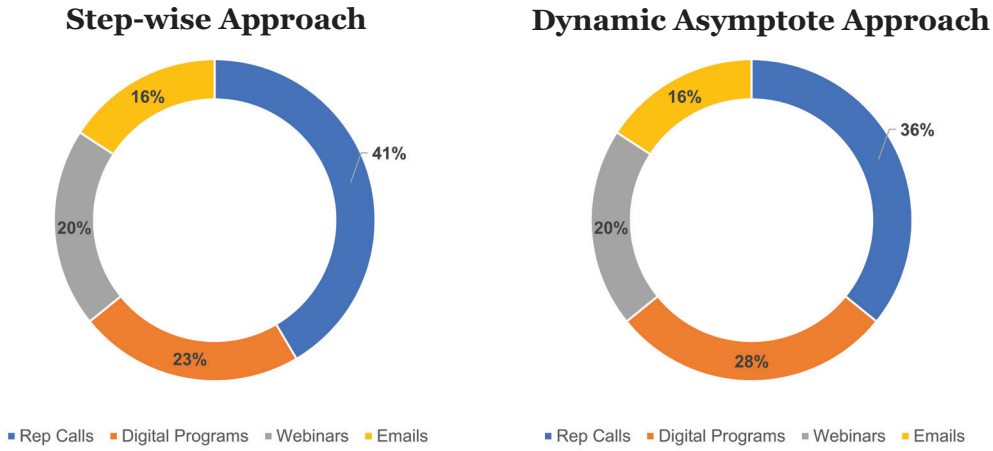


Figure 6 illustrates the split of HCP digital ads effect between direct and indirect effect.

2.3.3. Sales Attribution Among Individual Channels

Dynamic asymptote approach provides a truer distribution of the effects across channels compared to prior approaches, since it accurately allocates the sales effect to the channel actually causing it. Consider a case of DTC promotion channels targeting patients and increasing the patient traffic to a doctor. Now, when the physician is targeted successfully through the physician level promotion tactics (for example, rep calls and samples), the increased patient flow can be converted to NRx. This conversion should be attributed to the DTC channels that increased the patient flow. For example, when the PS campaign is successful in targeting the patients, the patients will learn about the options available to them and will visit the doctor's office. The NRx generated from these patients should be attributed to the PS campaign, instead of rep calls. The dynamic asymptote approach is successful in accurately capturing this effect.

Figure 7 represents the sales attribution to promotion channels with and without dynamic asymptote. Please note, a portion of total sales was also attributed to brand equity of the drug and previous writing behavior of the physician.

The difference in attribution is due to channel interaction. Rep calls show an effect of 41% in the prior approach but decreases to 36% in the dynamic asymptote approach. This difference of

Table 3: Comparison of Step-wise Asymptote and Dynamic Asymptote Approach

Spend			Step-wise Approach			Dynamic Asymptote Approach		
PS	SM	Total	Asymptote of PS	Effect	ROI	Asymptote of PS	Effect	ROI
$0	$1,000,000	$1,000,000	$700,000	$428,097	43%	$662,225	$567,787	57%
$200,000	$800,000	$1,000,000	$700,000	$719,880	72%	$647,281	$727,337	73%
$600,000	$400,000	$1,000,000	$650,000	$841,347	84%	$597,317	$883,596	88%
$700,000	$300,000	$1,000,000	$500,000	$774,595	77%	$578,694	$885,340	89%
$800,000	$200,000	$1,000,000	$500,000	$737,555	74%	$556,694	$669,912	67%
$1,000,000	$0	$1,000,000	$500,000	$482,163	48%	$500,000	$480,254	48%

ROI = Revenue / Cost

5% is coming from interaction components. This will lead to a more accurate ROI assessment of the channels, and a better budgeting decision. The final output of the whole exercise will be the promotion budget across channels for multiple scenarios. The scenarios could range from utilizing the historical budget optimally to identifying the budget that maximizes profit.

2.4 Limitations of Dynamic Asymptote Approach

While the approach overcomes challenges of prior approaches, it also has certain limitations of its own, as listed below.

- Some digital campaigns are run for short durations and hence do not have a lot of modeling data points. Due to a lack of enough data points, the model might not throw statistically significant results for those channels.
- The results might be difficult to interpret and visualize in the case of multiple interaction terms.
- Presence of too many parameters (Asymptote and Curvature for all interaction terms and the channel response curves) might make it difficult to create a model with acceptable p-values.
- The approach requires a theoretical model design as the first step. This involves talking to sales and marketing teams at the client organization, checking existing model design, using prior experience in running such models, and reviewing the existing literature to decide which channels interact with each other.

3. Strategic Business Benefits of Measuring Channel Interaction Using Dynamic Asymptote Approach

Measuring interaction using the dynamic asymptote approach lends numerous strategic benefits in marketing budget estimation.

3.1 Better Marketing Decisions

Since the dynamic asymptote approach better captured interaction effects, the sales attribution to promotion channels was more accurate. This enables better allocation of marketing budget. Table 3 lists different allocations for a total budget of $1M across SM and PS campaigns along with their corresponding effects. These effects were captured using two approaches – step-wise and dynamic asymptote. The highlighted row represents the best allocation in both the scenarios.

Step-wise approach provides a range of investment in one of the 2 interacting channels, in this case, SM. Since, the approach

Table 4. Comparison of Total Effect of TV and Print for Different Timing Scenarios

	Effect of rep calls (in MM)	Effect of HCP digital (in MM)	Effect from interaction (in MM)	Total Effect (in MM)
Only DTC TV is active	$17.067	$-	$-	$17.067
Only DTC Print is active	$-	$0.379	$-	$0.379
DTC TV and DTC Print active separately	$17.067	$0.379	$-	$17.447
DTC TV and DTC Print active within the shelf life of DTC TV	$17.067	$0.379	$0.240	$17.686

determines interaction effect within step-sizes was $200k, it is not possible to differentiate between interaction effect of a $200k spend in SM versus a $300k investment in SM (note that the asymptote of PS is the same in the range of $200k-$400k investment in SM). Thus, the step-wise approach suggests an investment of $400K in SM (maximum ROI of 84%). However, the dynamic asymptote approach enables evaluating interaction effect in this range more granularly (the PS asymptote changes based on the level of investment mix). This makes it clear that spending $300k in SM will result in a higher ROI compared to either $200k or $400k.

Thus, by capturing the interaction effect, spend allocation across channels is more effective. This will help achieve better effect and ROI for the same level of total investment.

3.2 Better Timing of Promotional Activity

Since promotional campaigns interact with each other, their effect depends upon when they are run. For example, consider a case where consumers are exposed to DTC TV and DTC Print for promotion of a drug. Since the two channels interact with each other, it makes sense to run the DTC Print promotion when the effect of the DTC TV promotion is on (i.e., during TV shelf-life). In this case, TV promotion has a shelf-life of five weeks. Similarly, the shelf-life of DTC Print is four weeks. If the DTC Print campaign is run within the shelf-life of the DTC TV campaign, the effect of TV will be represented by the equation below.

[6] *DTC TV Effect* = $[\{A_{TV\_Direct} + A_{Int}*\{1\text{-}EXP(-C_{Int}*(\text{Print GRP}))\}\}* \{1\text{-}EXP(-C_{TV}*(\text{TV GRP}))\}]$, where

A_{TV_Direct} = Asymptote of TV effect when no DTC Print campaign is run during TV shelf-life

A_{Int} = Asymptote of the interaction term

C_{Int} = Curvature of the interaction term

GRP = Gross Rating Point (applicable for both TV and Print campaigns)

Using the above formula, four scenarios were studied, and sales effect in all of them were compared. Table 4 lists the total effect generated from DTC TV and DTC Print in different timing scenarios.

Spend on DTC TV was $10M and spend on DTC Print was $400K.

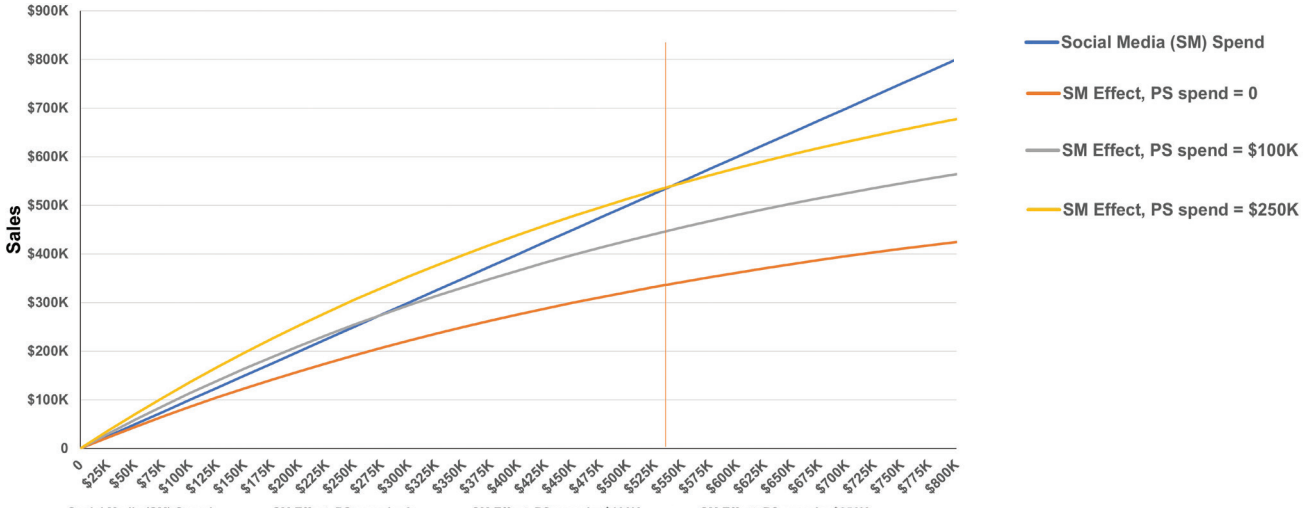
In this case, a marketer would prefer scenario 4.

Therefore, measuring interaction effect makes it possible for the marketer to make better promotion timing decisions.

3.3 Better Selection of Channels for Promotion

Marketers often find it difficult to explain the success of a few promotion channels, particularly for the channels which do not

Figure 8. Response Curves for SM at Different Levels of Investment in PS



directly affect sales in the immediate term.[9] The dynamic asymptote approach can help justify spend on these promotion channels by accurately capturing their effect on sales. Also, some of the channels generate more sales when they are switched on with other channels. In other words, the effectiveness (asymptote) of some channels increases when they are operated in proper combination. Take an example of SM promotion switched on with PS. Figure 8 evaluates the SM effect for various levels of investment in PS. It is seen that SM will be profitable till $500K spend, only if $250K is spent on a PS campaign. For any level of investment in PS below 250K, SM is not profitable. Clearly, the opportunity to earn an effect from SM would have been lost if interactions were not captured. Figure 8 represents three different response curves of SM (based on three levels of investment in PS) taken from a 3-D Cartesian-Coordinate-System that contains SM response curves for continuous levels of investment in PS and SM.

4. Conclusions

This paper tackles the problem of the changing landscape of pharma marketing in the emergence of newer digital channels. An analysis using interaction terms obtained accurate business insights from the regression model. The results were in line with the expectations and justified spending in digital channels. At the same time, the traditional approach was not able to justify the spending on those digital tactics.

Measuring interactions will become even more critical in the future since promotion in non-traditional channels will increase, thus, generating greater effects caused from their interaction. Further, this can assist pharma companies with automating their real-time campaign management solutions, which will enable them to make decisions on when and which campaign to run based on their effect. As the pharma marketing landscape continues to evolve, the objective of marketing will also change from sales focus to value focus. Promotion to other stakeholders (i.e., payers and patients) will become more sophisticated. This means that the dependent variable will not just be sales units or NRx, but also other success factors such as patient persistency and adherence. Further, some channels might be specifically targeted towards payers with an aim to better the drug coverage in their plans and

improve health and economic outcomes. This means that interaction happens not just among promotion channels to a single stakeholder, but also among promotion to different stakeholders. Therefore, measuring interaction will become increasingly critical for budgeting decisions.

Pharma marketing has been evolving over many years. Pharma practitioners will observe the onset of newer promotion tactics and ways to target different stakeholders. Hence, the methodology should also evolve. The analysis elaborated here was able to use a combination of negative exponential and linear transformations to build a regression model. Researchers can further use a few more transformations or techniques to adapt to these changes and increase accuracy of effect and ROI measurements.

About the Authors

***Shubham Lahoti**, an Associate Director at Axtria, has more than 12 years of experience in analytics, modeling, and technology projects with more than 7 years in sales and marketing optimization in the pharma industry. He has led several analytics projects that involve segmentation, multi-channel promotional analysis, optimization, targeting, patient analytics, predictive modeling, etc. Shubham holds a Bachelor's degree from the Indian Institute of Technology, Bombay.*

***Varun Jain** is an Associate Director at Axtria where he works in the Decision Sciences practice. He works on marketing analytics projects and technology solutions for top pharma clients. He has led large marketing-mix and budget optimization assignments across various therapeutic areas such as vaccines, oncology, and rare diseases. His focus has been on running analytics for various kinds of marketing problems across traditional and digital channels. Varun holds a Master in Business Administration from the Indian School of Business (Hyderabad), with a major in Finance and Strategy, and a Bachelor's in Engineering from the Information Technology from Netaji Subhas Institute of Technology, Delhi University.*

***Adarsh Gautam** is a Project Lead at Axtria where he works in the Decision Science practice. In a very short span of time, he has worked with multiple pharmaceutical clients and helped them in their marketing and operations decisions by bringing cutting edge-data analytics and technical expertise. In addition to this, he has also worked on implementing Axtria's MarketingIQ™ platform across various life science organizations. He has also led several Marketing Mix and Sales Force Sizing projects during his stint at Axtria. Adarsh holds a Master of Technology degree in Biomedical Engineering and a Bachelor's in Bio-engineering from the Indian Institute of Technology, BHU (Varanasi).*

References

1 Chressanthis G and Naphade K. Is it time to adjust the pharma PDE sales force optimization model? *Axtria Research Hub*, published online April 2017, available at http://insights.axtria.com/whitepaper-is-it-time-to-adjust-the-pharma-pde-sales-force-optimization-model.

2 Paul J. Why you should adopt a cross-channel marketing strategy. *Candidsky*, published online 20 Feb 2018, available at https://www.candidsky.com/blog/adopt-cross-channel-marketing-strategy/.

3 Evolving pharmaceutical marketing strategies: digital as a Key Component of Multichannel Marketing, *GBI Research*, Jan 2017, available at http://gbiresearch.com/report-store/market-reports/cbr-pharma-insights/evolving-pharmaceutical-marketing-strategies-digital-as-a-key-component-of-multichannel-marketing.

4 Prasad A, Kalyan R and Russell S. Planning marketing-mix strategies in the presence of interaction effects. *The Institute for Operations Research and the Management Sciences* 2015; 24: 25-34.

5 Pearl J. *Causality: Models, Reasoning and Inference*. 2nd ed. New York, NY: Cambridge University Press; 2009. Chapter 5, Causality and structured models in social science and economics, pp. 133-140.

6 Wood D and Lahoti S. Optimizing account managers and sales reps. Pharmaceutical Management Science Association Annual Conference, 1 May 2018, San Antonio.

7 Pinheiro J, Bates D, Debroy S, et al. R-package. NLME, SA and GENSA (version 3.5.0).

8 Sievert C, Parmer C, Hocking T, et al. R-packages. PLOTLY and DATA.TABLE (version 3.5.0).

9 Wood L. Evolving pharmaceutical marketing strategies 2017: digital trends are changing marketing strategies within the pharmaceutical field – Research and Markets, *Business wire*, published online 2 Mar 2017, available at https://www.businesswire.com/news/home/20170302005998/en/Evolving-Pharmaceutical-Marketing-Strategies-2017-Digital-Trends.