

# **View from the ‘Driver’s Seat’**

**New Rules of the Road for Developing  
and Applying Customer Behavior Models**

**PHARMA MARKET RESEARCH CONFERENCE**

**February 6, 2020**



**NAXION**  
RESEARCH > CONSULTING

# Driver Modeling Defined

Family of *inferential* techniques used to explain customer actions by identifying and quantifying predictors



What drives use/affinity for a *single* brand?

What drives use/affinity for one brand *over another*?

# Rationale and Assumptions



**Customers are not always self-aware or fully candid**

**Statistical relationships between predictors and outcome measure(s) can offer critical insights not otherwise available**

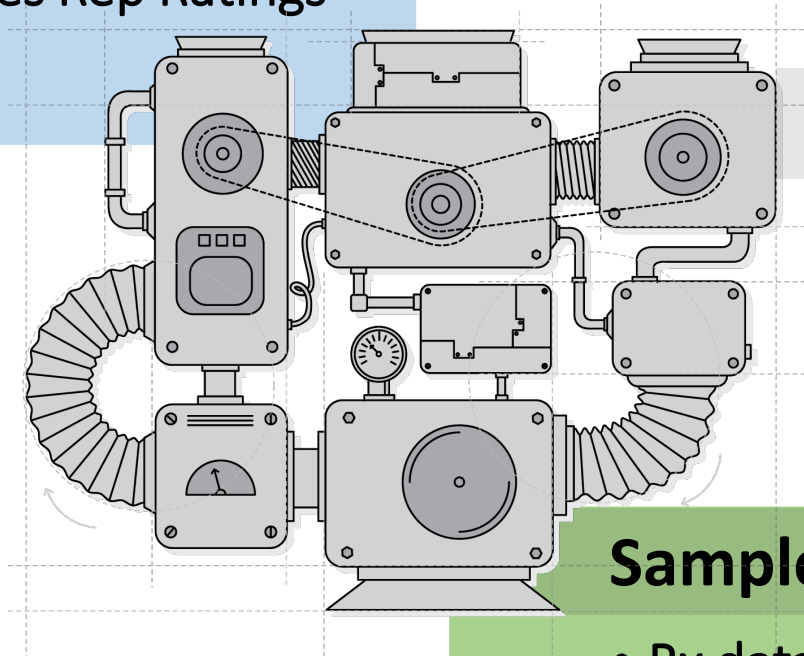
**The relationships between predictors and outcomes are causal**

**Despite noise, inferences from modeling may be more trustworthy than what people tell us**

# Elements of a Driver Model

## Sample Inputs

Product Performance Ratings  
Manufacturers & Sales Rep Ratings  
MD 'Demos'



## Predictive Model

## Sample Outcomes

- Rx data
- Self-reported Rx
- Brand Affinity Metric

# Choosing a Method: The Basics

## PROS

- Requires less data
- Easily interpreted

### Univariate Correlation

## CONS

- No integration
- No equation
- Does not address intercorrelations

- Provides formal model to estimate impact of change in predictors
- Handles multiple variables in one model

### Stepwise Regression

- Requires more sample
- Automated eliminates of correlated variables

# More Advanced Options

## PROS

## CONS

- ✓ Reduces model bias
- ✓ Finds non-linear relationships
- ✓ Handles multi-collinearity

### Random Forest

- ✓ More labor intensive
- ✓ Can't estimate impact of changes in predictors

- ✓ Reduces model bias
- ✓ Provides equation to estimate impact of changing predictors
- ✓ Handles multi-collinearity

### Shapley Regression

- ✓ More labor intensive
- ✓ Only for continuous outcomes

- ✓ Complex “mother of all models”
- ✓ Handles multi-collinearity

### Structural Equation

- ✓ Very data and labor-intensive
- ✓ iROI rarely justified

**So, You've Got Data  
and Statistics at  
Your Disposal...**



**...What Could  
Possibly Go Wrong?**

# Small Sample Size

Many MD survey datasets are not large enough to model all the attributes, resulting in overfit models



Rule of thumb: sample size  $\geq 10x$   
number of predictor variables

If  $n = 50$ , maximum number of input  
*variables for reliable model is 5*

Consider using univariate  
correlations and marketing  
judgement to eliminate variables

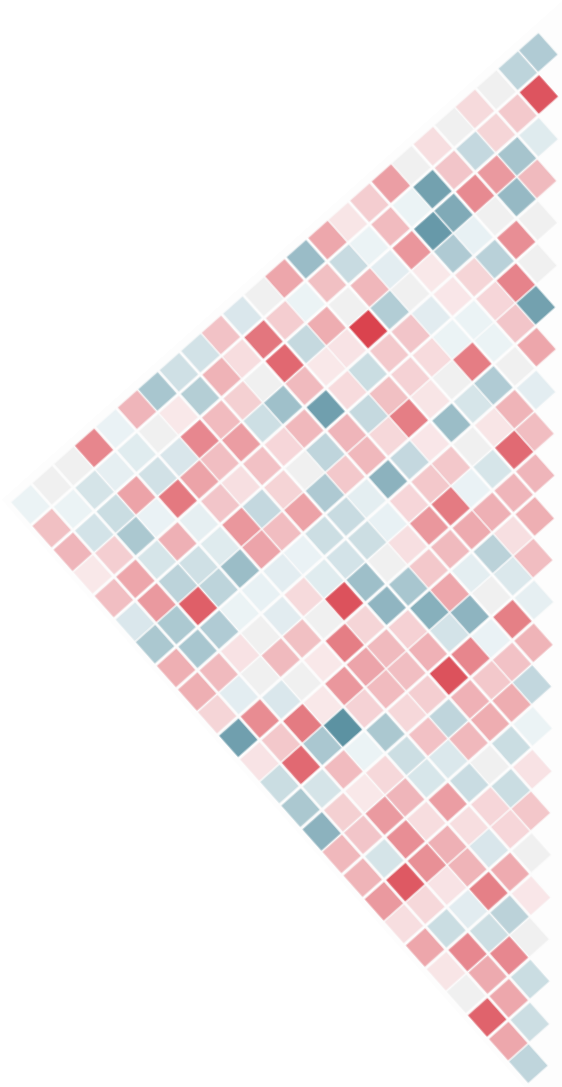




# Multi-collinearity

Attributes are frequently correlated with each other – particularly when specific types of efficacy, safety, or tolerability attributes are proliferated

*Consider relying on more advanced methods (e.g., Random Forest and Shapley Regression)*



# Insufficient Variability on Attributes



Infrequent usage opportunities limit inherent variability of what you aim to predict

*Consider ratings as outcome*

Strong consensus about product performance on an attribute limits its ability to predict

*Caution: Lack of predictive power does not = irrelevance*

# Face Validity and Contradictions

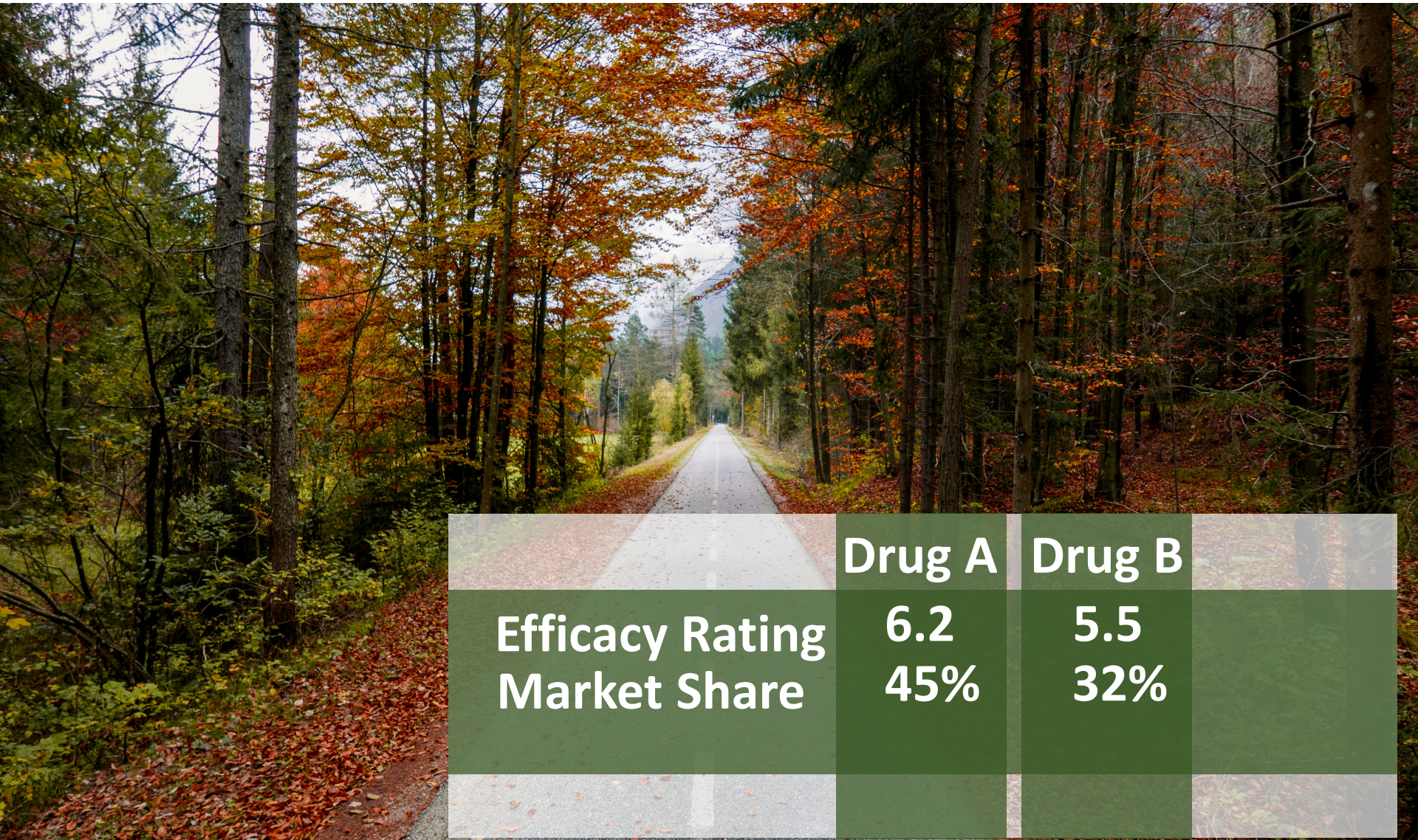


Results that contradict  
stated importance  
subject to question

*Consider how data have  
been presented, role of  
statistical significance,  
and whether a broader  
view might help*



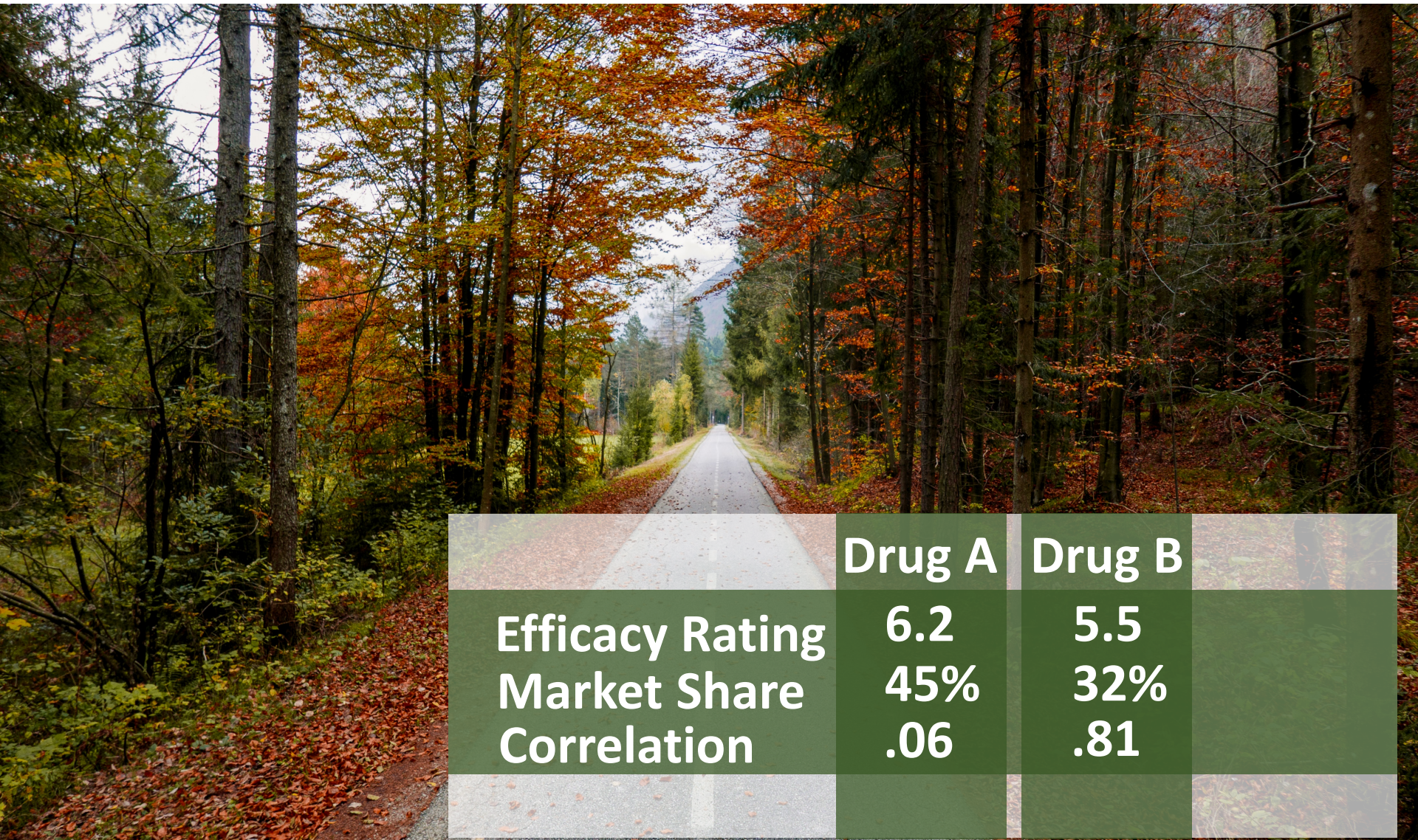
# Market Coherence: Value of Integrated Model



	Drug A	Drug B	
Efficacy Rating	6.2	5.5	
Market Share	45%	32%	



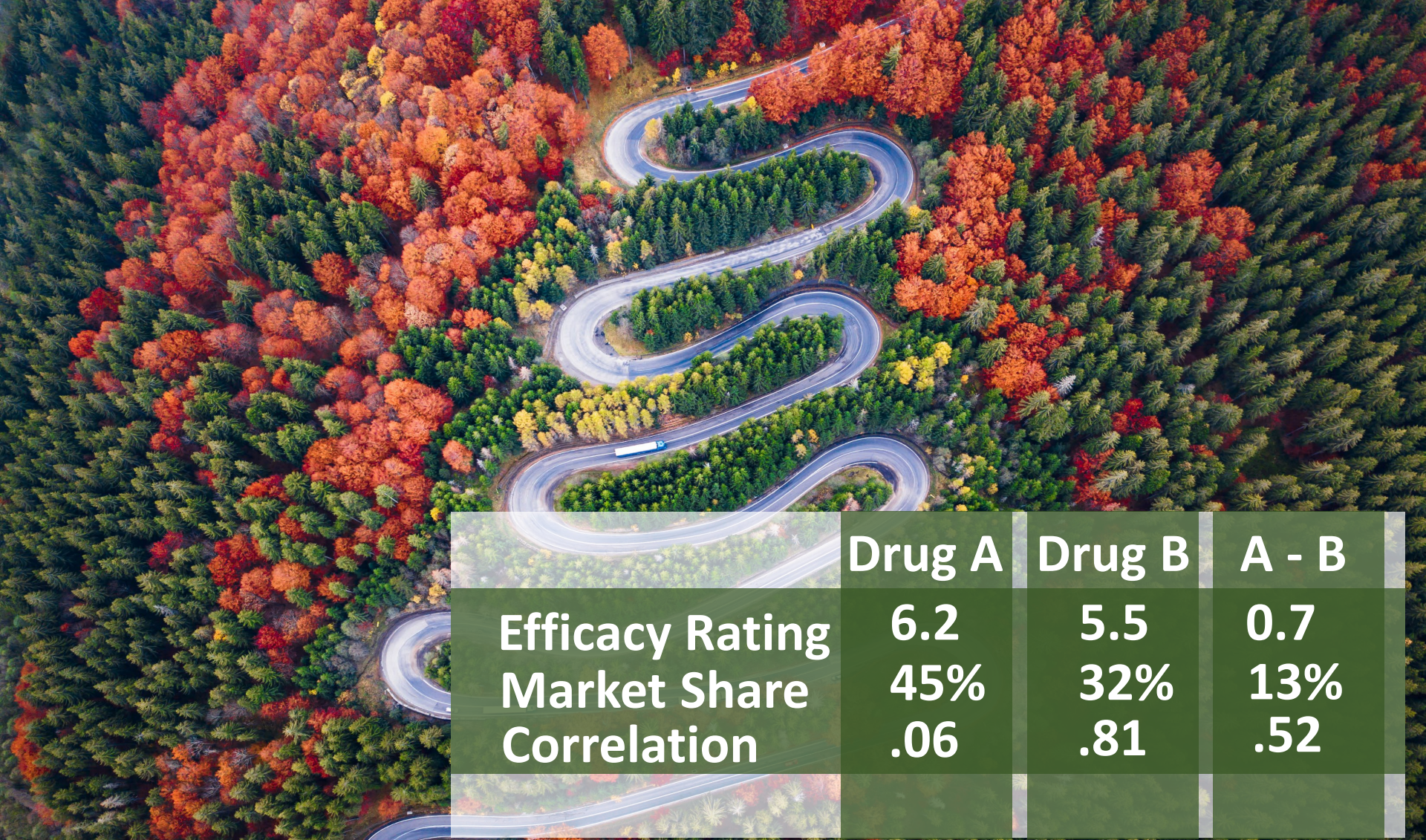
# Market Coherence: Value of Integrated Model



	Drug A	Drug B	
Efficacy Rating	6.2	5.5	
Market Share	45%	32%	
Correlation	.06	.81	



# Market Coherence: Value of Integrated Model



	Drug A	Drug B	A - B
Efficacy Rating	6.2	5.5	0.7
Market Share	45%	32%	13%
Correlation	.06	.81	.52

***Caveat: Difference measures can limit variability***





# Methods in Sum

## Criteria

Correlation/  
Univariate  
Regression

Step-Wise  
Regression

Random  
Forests

Shapley  
Regressions

Provides a formal model		✓		✓
Ranks all attributes	✓		✓	✓
Handles multi-collinearity		✓	✓	✓
Provides statistical significance	✓	✓		
Identifies non-linear relationships			✓	
Handles binary outcomes	✓	✓	✓	
Minimizes model overfit			✓	✓



# Rules for Safe Driving



**Start by being inclusive and creative**



**Feature-engineer to strengthen model**



**Be prepared for iteration**



**Prune empirically to improve predictions and give access to broader range of analytic tools**



**You need someone at the wheel**

# You Still Need Someone at the Wheel





**Michael Polster, Ph.D.**  
*Senior Vice President*

**NAXION**  
RESEARCH > CONSULTING