Machine Learning for Humans with Humans

Nasir Bhanpuri, PhD
Clinical Informatics Data Scientist, Virta Health

Great Examples of Automated Machine Learning

Automatic colorization of Black and White photo

Lizuka et al 2016

Instant translation between languages with smart phone app

Goolge Translate, Hadod 2017

Requires

- A lot of data
- Some degree of automated errors are acceptable
- Limited interpretability

Humans + machines > machines?

Humans + machines >> humans

Lunit
INSIGHT TM 2018

Shabazz, 2007

Anson Williams
Freestyle Chess Champion
“Centaur”

Tumor detection, human + algorithm more accurate than radiologist alone and very helpful for non-radiology physicians

Modeling Spectrum

More Abstract

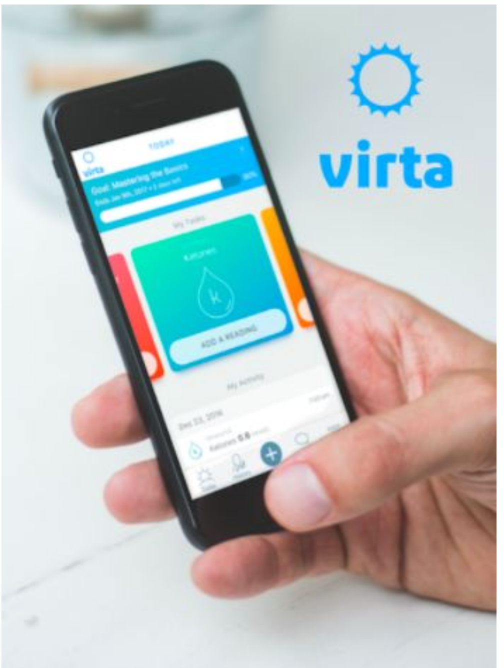
- Flexible Algorithms
- Learn parameters
- Less interpretable
- Focus on accuracy & application

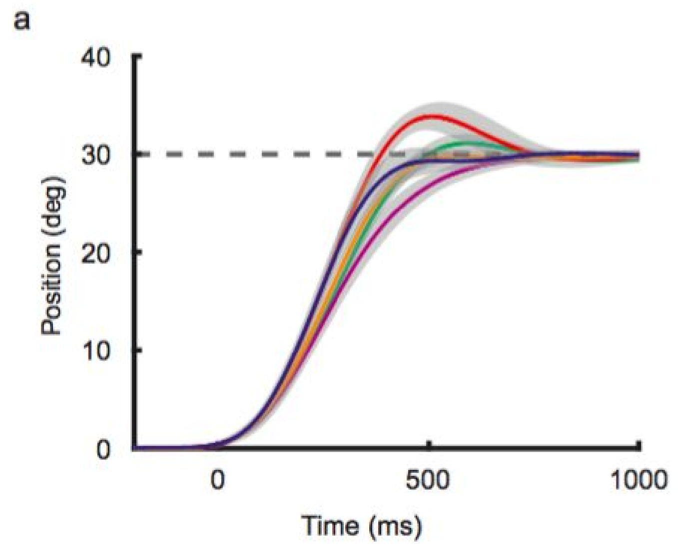
Less Abstract

- Underlying mechanisms
- Fit parameters
- More interpretable
- Focus on insights & discovery

Automatic colorization of Black and White photo

SUPER BOWL





Lizuka et al 2016

Bhanpuri et al 2014

Why football, health, music, and data?

“What makes him successful is the way that he analyzes information. He is not just hunting for patterns. Instead, Bob combines his knowledge of statistics with his knowledge of basketball in order to identify meaningful *relationships* in the data.”

Nate Silver, Statistician on Haralabos “Bob” Voulgaris, Sports Bettor

Nasir Bhanpuri

Predicting NFL Winners with NFL.com Writers

Playoffs and Super Bowl

Super Bowl!

Most watched television broadcasts in the United States

No.	Show	Viewership (millions)	Date
1	Super Bowl XLIX	114.4	February 1, 2015
2	Super Bowl XLVIII	112.2	February 2, 2014
3	Super Bowl 50	111.9	February 7, 2016
4	Super Bowl LI	111.3	February 5, 2017
5	Super Bowl XLVI	111.3	February 5, 2012
6	Super Bowl XLV	111.0	February 6, 2011
7	Super Bowl XLVII	108.7	February 3, 2013
8	Super Bowl XLIV	106.5	February 7, 2010
9	*M*A*S*H* (Finale)	105.9	February 28, 1983
10	Super Bowl XLIII	98.7	February 1, 2009

wikipedia.org

Why is football so popular?

Variability!*

*(And fun halftime show and commercials)

Playoff Bracket, 12 Teams → 2

Over last 6 years, team with better seed (or record) won 65% of games

Can we do better than that?

short answer is:

- not definitively given small sample, but trending in right direction
- Some suggested insights, but requires more observations

Data used for training & testing

2004—2011

2012—2014...

Training, cross-validation

Testing

Nfl.com
(Bhanpuri, 2016-2018)

Nasir Bhanpuri

Machine Learning for Insights (and Decision-Making)

- **Goal:** Predict playoff winners
- **Questions:** Who will win each game? What does/does not matter?
- **Users:** Mostly fun (writers, players, coaches, GMs)
- **Collaborators:** NFL.com writers
- **Model Requirements:**
  1. Accuracy
  2. Insights on factors
- **Model Benefits:** Predict winners, understand strengths and weaknesses better

Tom Blair

Nfl.com
(Bhanpuri, 2016-2018)

Feature Selection & Algorithm Selection

Based on football knowledge

- 110 Features (Characteristics, Factors)
- First 16 games of season
- Feature importance reduce to **Top 37**
  - Univariate ROC + experimentation w/ different feature combos
- SVM (V2.0) [linear kernel]
  - R: caret + kernlab (Platt scaling)
- + Linear Regression (V2.0m)
  - 69% Accuracy
    - 95% CI ~±11%
    - Current sample size cannot tell if actually better than "Top Seed" approach or just lucky

Offense
- Expected Pass Pts
- Pass Yds/attempt
- ...Rush TD

Defense
- Turnovers
- Passing TD allowed
- Points Allowed
- ...Rush TD allowed

Special Teams
- Punt Return Yards (-)
- *Not FG*

Overall
- Wins
- Strength of schedule

profootballreference.com

Tom Blair

Model Results: Insights—Feature importance & Directionality (possibly*)

Increase Strength	Decrease Strength	Less Important
Passing statistics	Points Allowed	Field goals made
Rushing statistics	Touchdowns allowed	Team
Turnovers	Penalty yards	
Strength of schedule	Punt return yards	

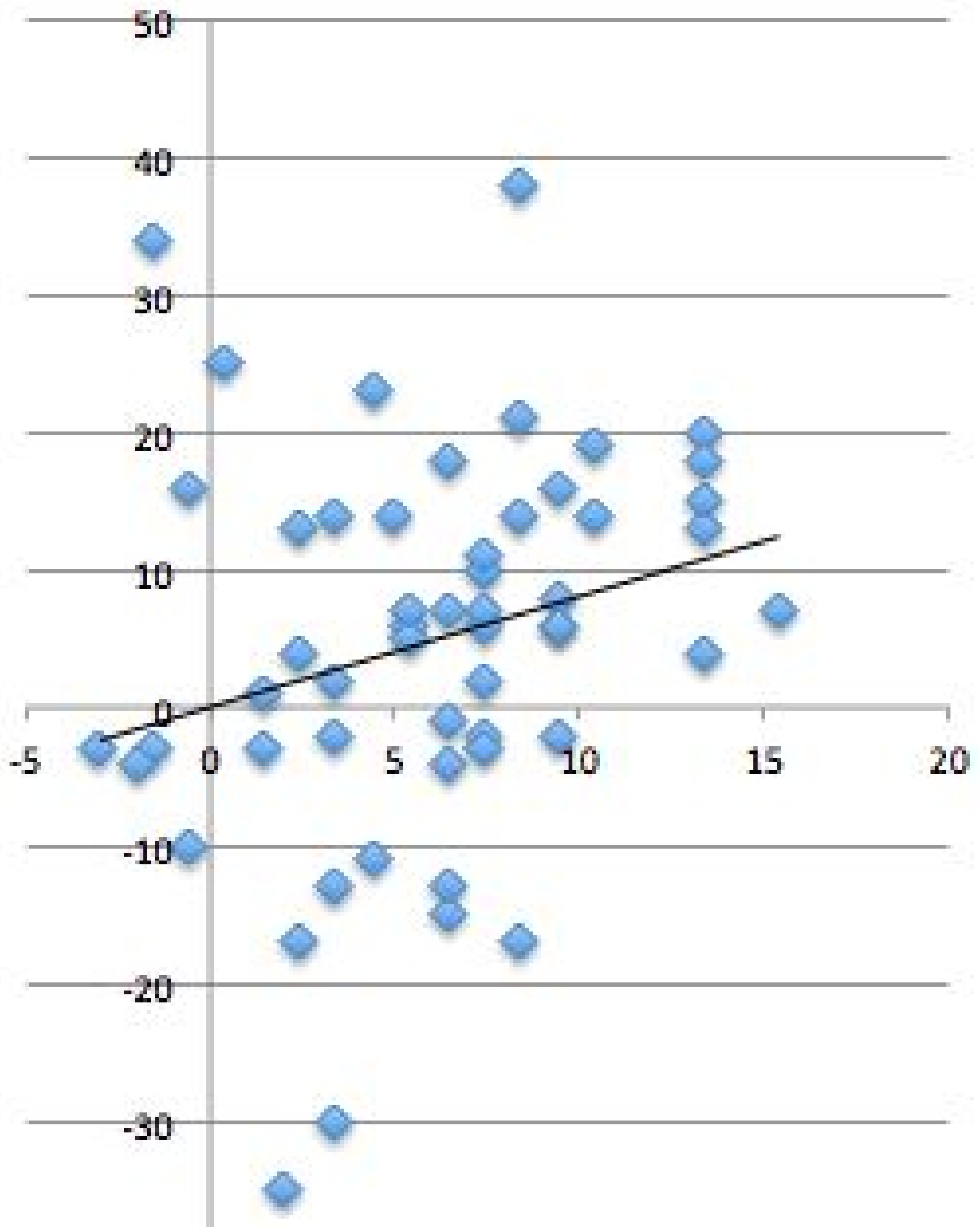
*These are "suggestive" insights from univariate analysis, may differ from final model

More hypothesis generation than conclusive

Experimentation could help confirm, though not always practical

Model Results

2017 Super Bowl Rating (V 2.0)	
New England Patriots	21.9
Kansas City Chiefs	18.5
Minnesota Vikings	17.4
Pittsburgh Steelers	17.3
New Orleans Saints	17.2
Los Angeles Rams	16.7
Philadelphia Eagles	16.4
Carolina Panthers	16.0
Jacksonville Jaguars	14.4
Atlanta Falcons	14.3
Buffalo Bills	12.9
Tennessee Titans	12.5



- $P < 0.05$
- High variance
- Majority correct over large sample but low confidence for any single game

Suggested insights (aka potential impact)

- "Hot streak" doesn't matter
- "Elite" QBs show up in the stats
- Passing success more important than rushing
- Special Teams don't matter as much...
  - Punt returns (unreliable) & FG needs field position?
- *Important caveats*
  - Late season injuries unaccounted for (Shazier, Gronkowski)
    - human + machine
  - Relatively small validation data and large confidence interval, all luck?!

Predicted Winner	Predicted Margin of Victory	% Chance of winning
NE	4–5	63%

Estimated Confidence Interval: 52–74%

How might Eagles win?

- Aggressive play, take chances
- Patriots turnover(s), unlikely
- Key Patriots injury (never hope for this, but reality)

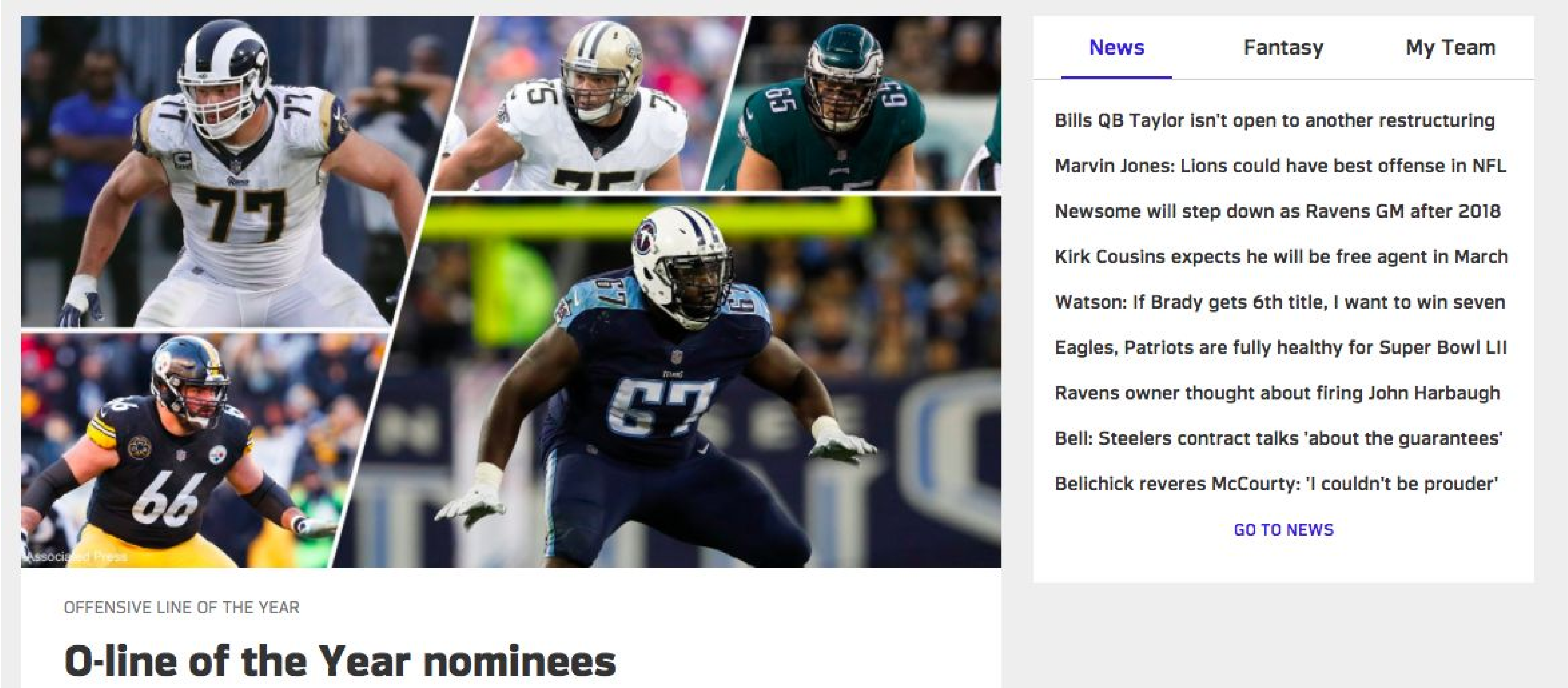
V3

- Team dependent home/away performance
- Adjustment for player injuries

Nfl.com
(Bhanpuri, 2016-2018)

Thanks NFL.com writers/editors!

More than news: Analysis, Research
Sidelines (long form), Oklahoma Drill (Interviews), and more...

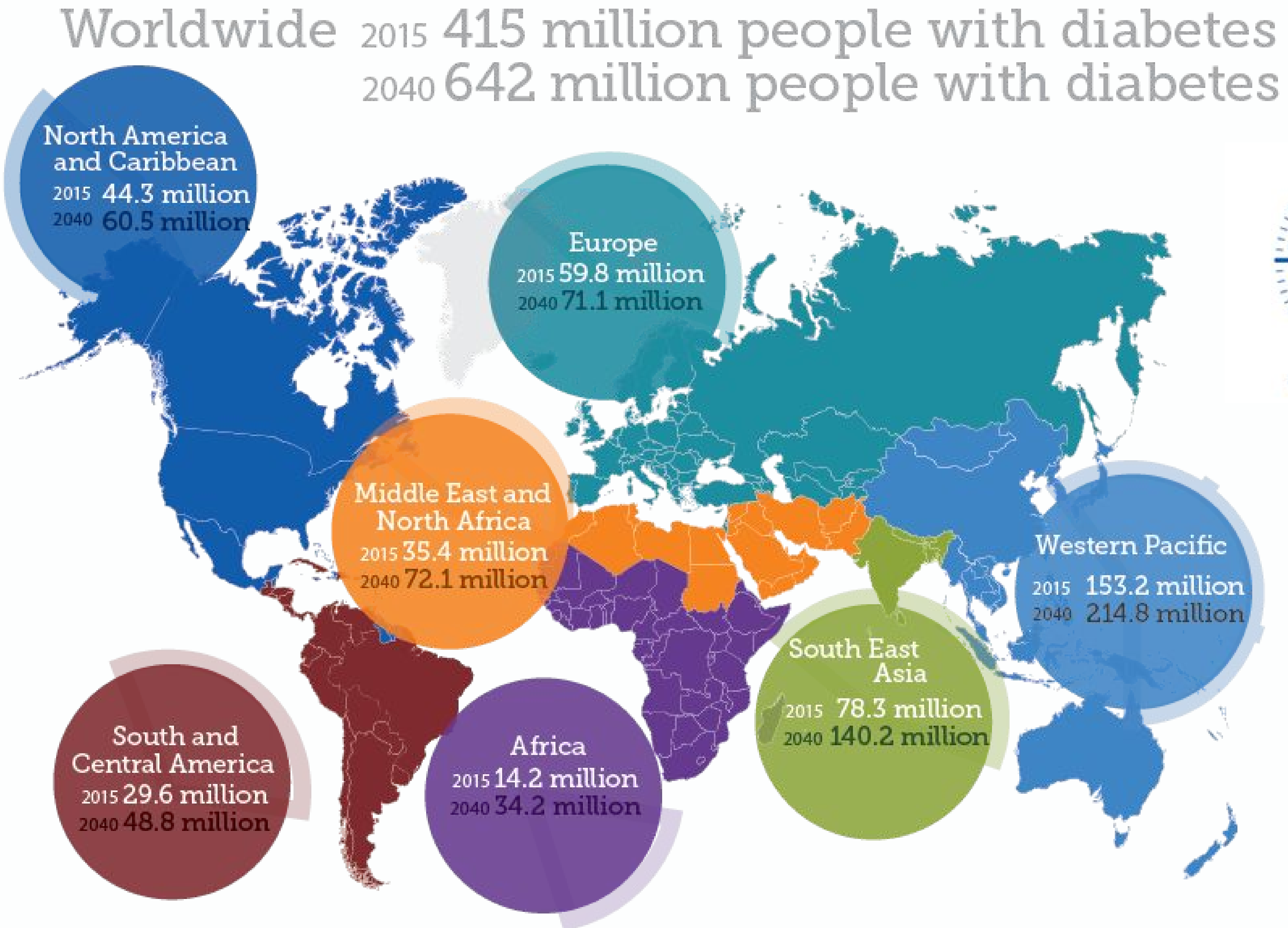


Nasir Bhanpuri

Improvements in Patient Retention with Virta Health Coaches

Continuing Virta treatment for at least one year

Diabetes is a global, accelerating, expensive!

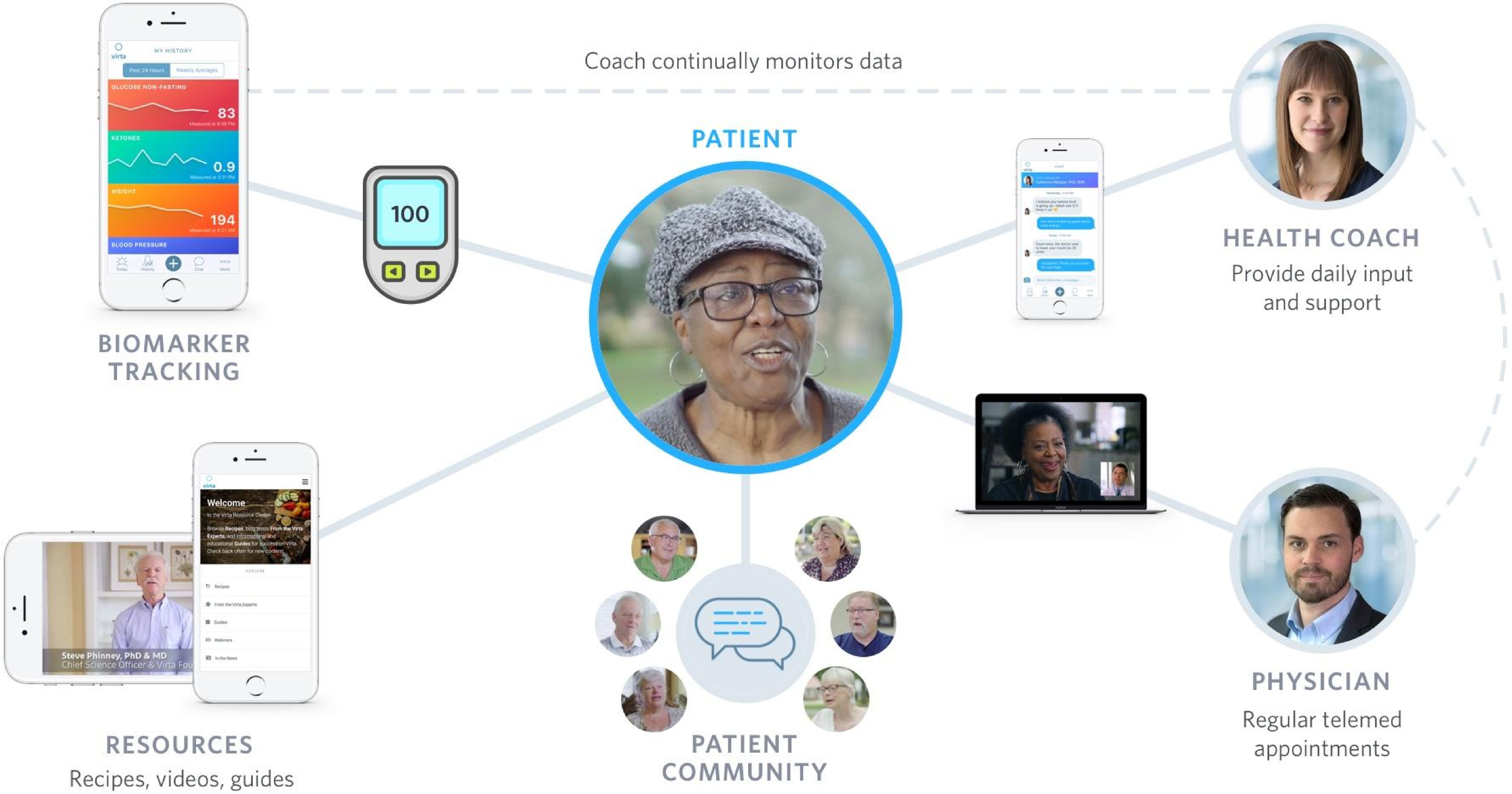


virta

Reverse Diabetes in **100M by 2025**

(International Diabetes Federation, 2015)
https://www.idf.org/about-diabetes/what-is-diabetes

Online Type 2 Diabetes Reversal Clinic



Nasir Bhanpuri

Clinical Trial Results

Summary of 2-6 month outcomes

87% OF PATIENTS REDUCED OR ELIMINATED INSULIN [1]

56% OF PATIENTS REDUCED THEIR HBA1C BELOW DIABETIC LEVEL [1]

12% AVG WEIGHT LOSS AT 6 MONTHS [2]

1. McKenzie AL, Hallberg SJ, Creighton BC, Volk BM, Link ™, Abner MK, Glon RM, McCarter JP, Volek JS, Phinney SD. A Novel Intervention Including Individualized Nutritional Recommendations Reduces Hemoglobin A1c Level, Medication Use, and Weight in Type 2 Diabetes. JMIR Diabetes. 2017;2(1):e5
2. Preliminary 5 Month trial data.

Summary of 1 year outcomes

60% OF PATIENTS REVERSED THEIR TYPE 2 DIABETES

94% OF PATIENTS REDUCED OR ELIMINATED INSULIN

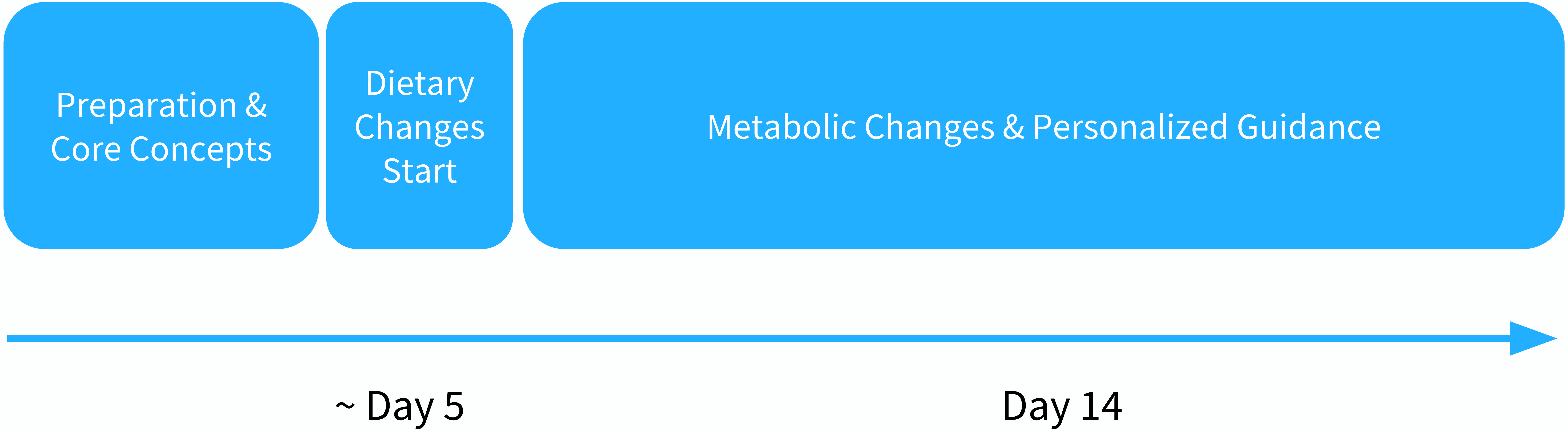
1.3% AVERAGE HBA1C REDUCTION AT ONE YEAR

30 lbs AVG WEIGHT LOSS AT ONE YEAR (12%)

3+ IMPROVED METABOLIC CONDITIONS

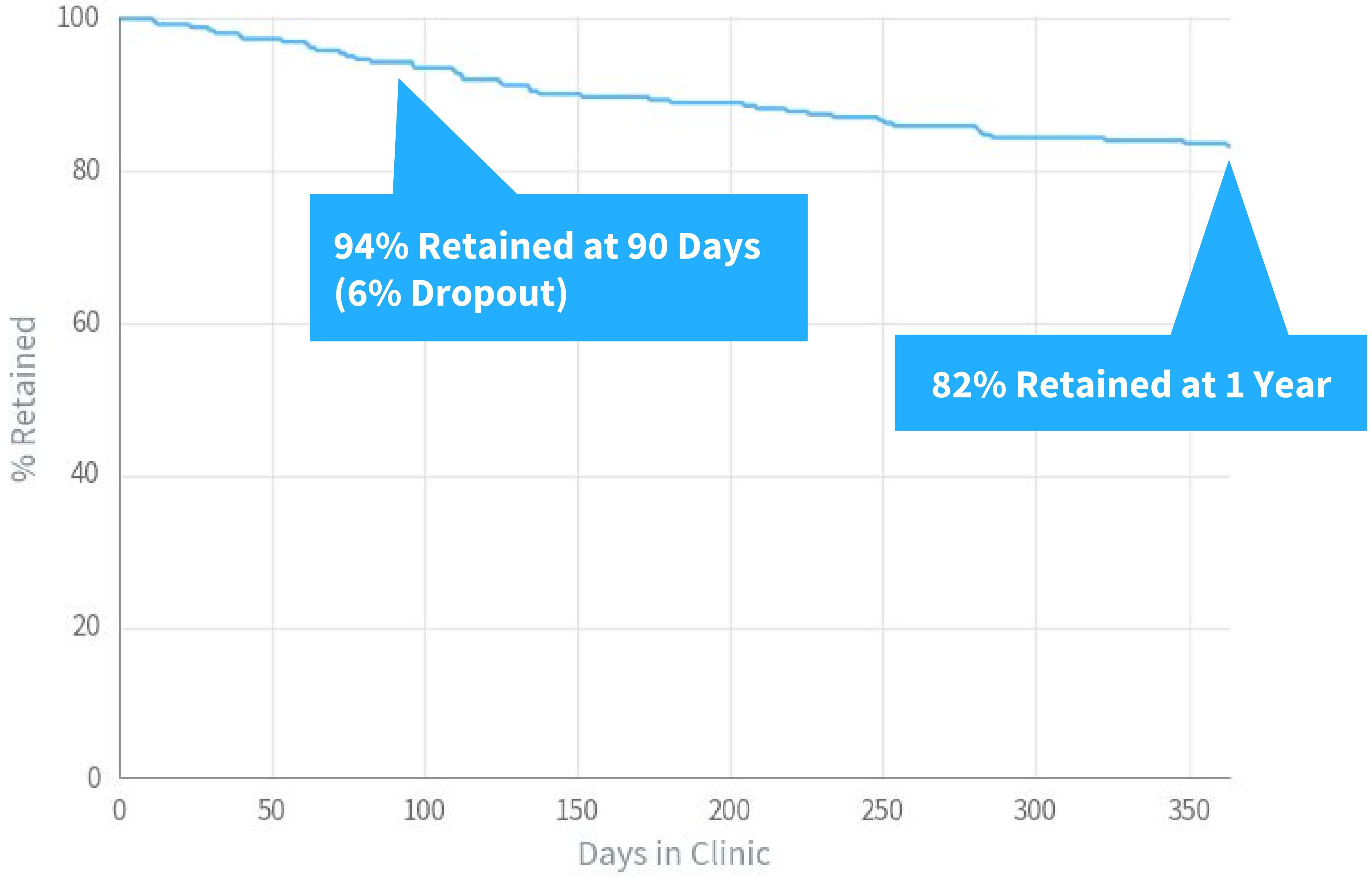
Hallberg SJ, McKenzie AL, Williams P, et al. Effectiveness and Safety of a Novel Care Model for the Management of Type 2 Diabetes at One Year: An Open Label, Non-Randomized, Controlled Study. Diabetes Ther. 2018. DOI: 10.1007/s13300-018-0373-9

Early in Patient Journey



Nasir Bhanpuri

High retention, but can improve

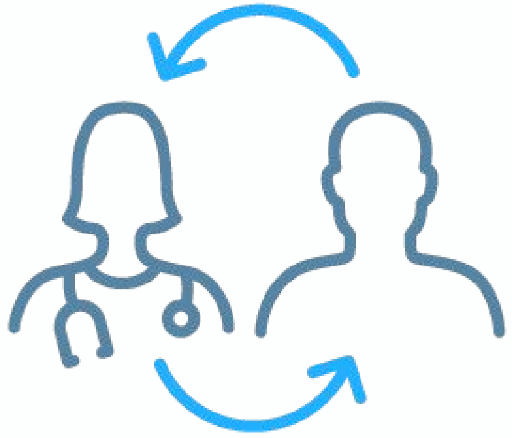


Source: IU Health Arnett - Virta Clinical Trial Data (Hallberg et al., 2017)
N = 158
*Those not “retained” either requested to terminate Virta services (usually because of unrelated health/family issues or undisclosed personal choice) or were removed from the study due to noncompliance and concerns related to safety.

Machine Learning to Drive Action and Decision-Making

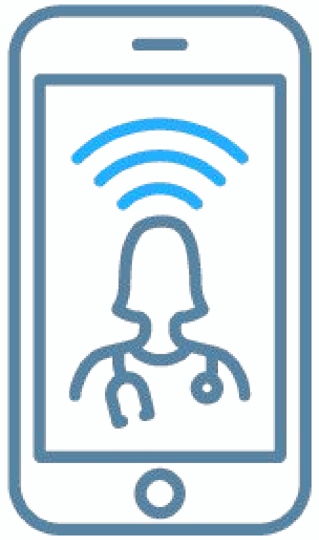
- **Goal:** Increase long-term retention rate of patients
- **Questions:** Who is at risk of dropping out? *Why* are they dropping out?
- **Users & Collaborators:** Clinicians
- **Model Requirements:**
  1. Easy to communicate to clinicians
  2. Accuracy
- **Model Benefits:** Prioritization and insight into underlying factors

Feature Selection & Algorithm Selection



Dedicated health coach

Text Messages
- Length
- Count/Freq
- Topic



App and biomarker tracking tools

App Data
- Weight
- Glucose
- Symptoms

Based on Clinical Data and Research

- 108 Features (Characteristics)
- First 14 days of data
- Feature selection for **Top 15**
  - Random Forest Out-of-bag error
- Logistic regression
- AUC = 0.78 (30% test set)

Model Results: Insights—Feature Directionality*

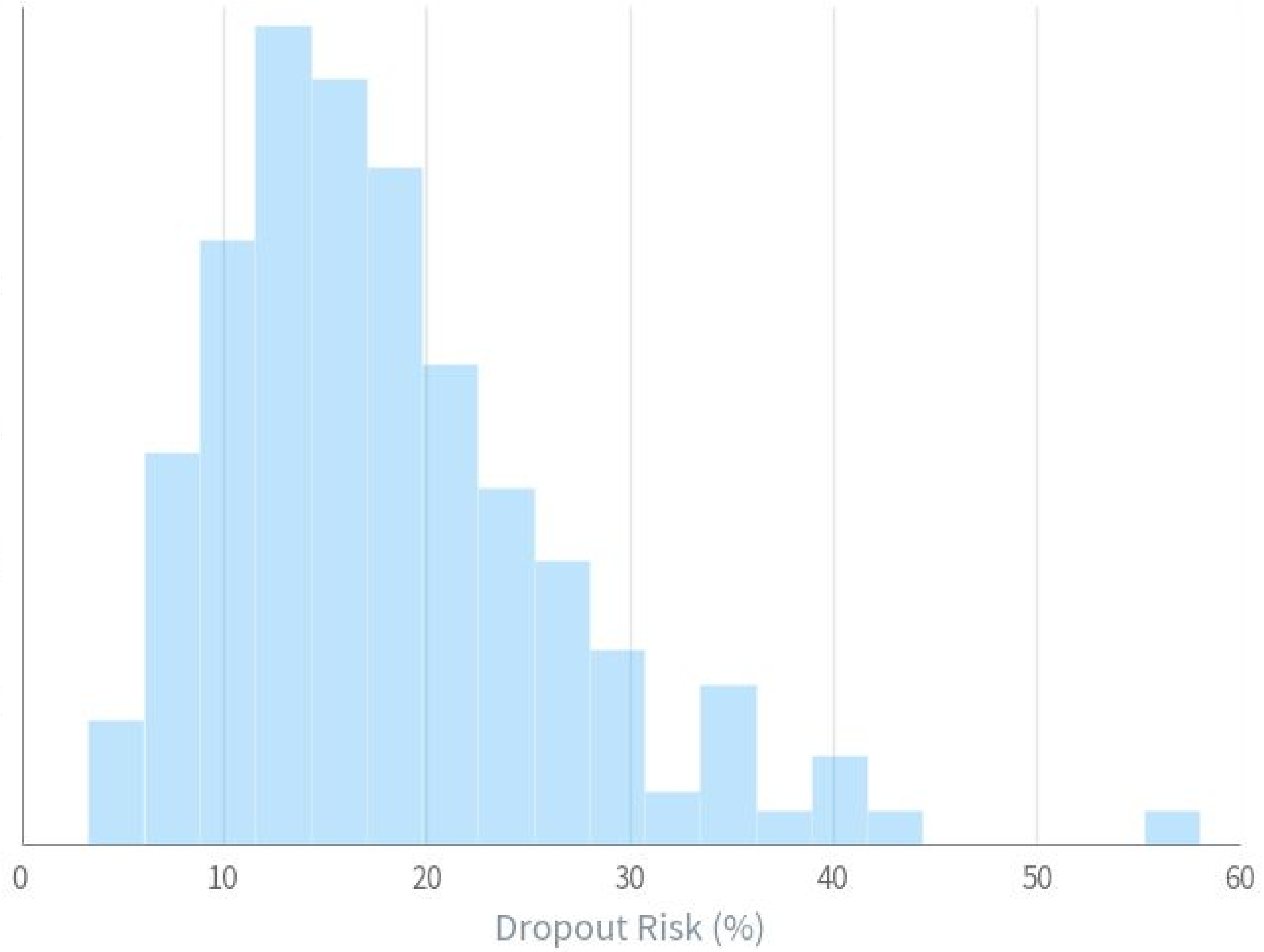
Increase Dropout Risk	Decrease Dropout Risk
Time to dietary change	Age
Texts about discomfort	Texts about challenges
Fatigue	Urgent texts
Opiate/Pain meds	Weight loss

*These are “suggestive” insights from the data and not definitive

More hypothesis generation than conclusive

Experimentation could help confirm, though not always practical

Model Results: Dropout Risk

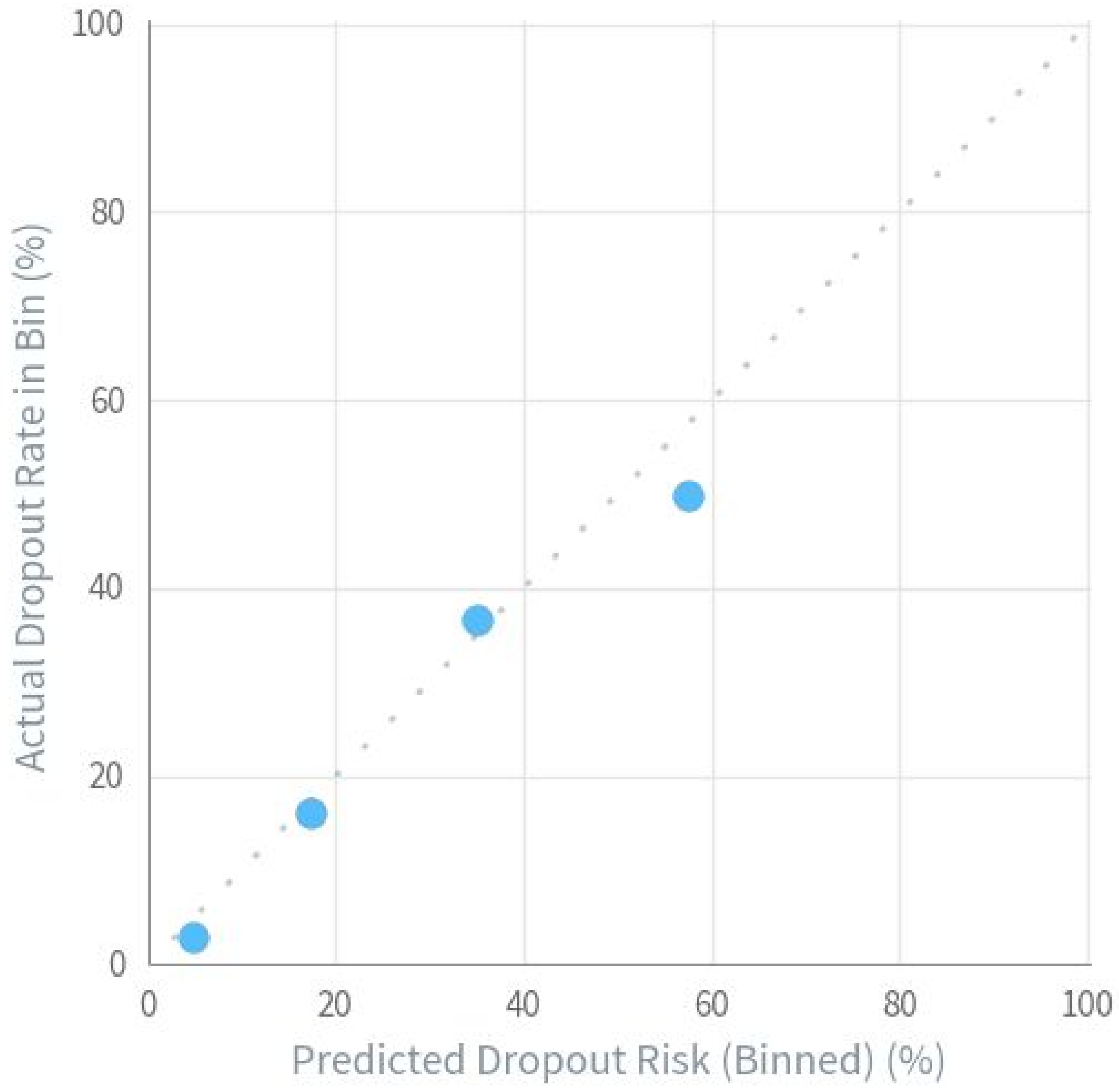


Distribution of Dropout Risk

- Average: 18%
- 50th percentile: 16%
- 90th percentile: 37%

Model Validation

- AUC: 0.78



Risk Level	Dropout Risk
Low	0 - 10%
Medium	10 - 25%
High	25 - 45%
Very High	45 - 70%
Extremely High	70 - 100%

\+ Anecdotal confirmation from coaches

Nasir Bhanpuri

Actionable Insights

Who?

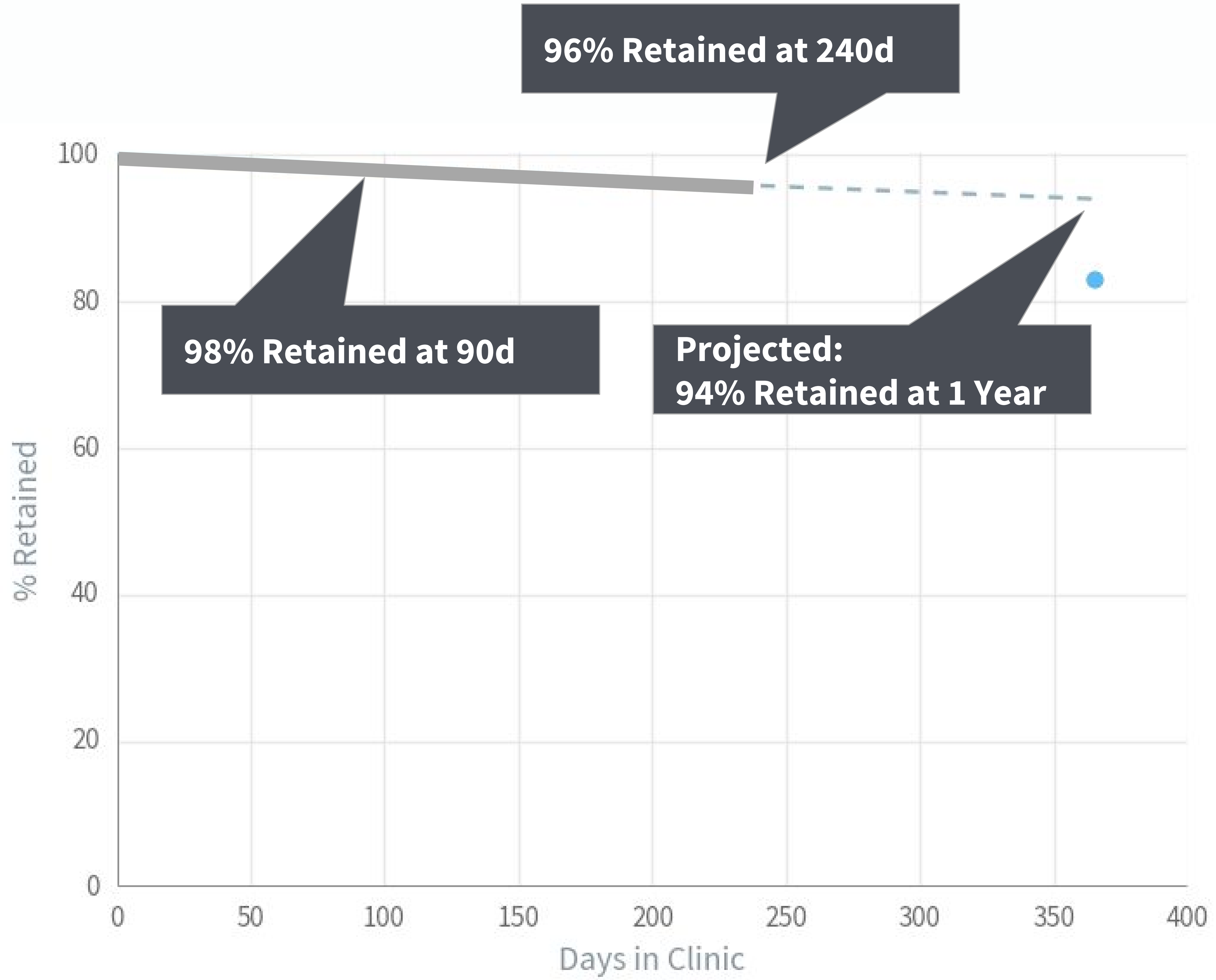
Patient ID	Dropout Risk
1	31.9 %
2	30.4 %
3	27.4 %

Who? Why? (What to do?)

Patient ID	Dropout Risk	Text Discomfort	Fatigue Count	Weight Change	...
2	High ~35%	9	5	-4.3	
3	High ~35%	14.5	0	-4.0	
1	High ~35%	9	2	-1.0	

Impact

- Coach impressions:
  - Prioritize additional outreach
  - Focus efforts
  - Human + machine
- Dropout rate down **66%**
- *Important caveats*
  - Different population
  - Evolving product



V2 & Other modeling efforts

- Daily prediction
- Retrain with different population
- Predict weight change, HbA1c change (blood sugar), etc.

Collaborators

James McCarter, MD, PhD
HEAD OF RESEARCH

Amy McKenzie, PhD
RESEARCH

Jackie Lee, PhD
DATA SCIENCE

Amit Shah
HEAD OF OPS & CUSTOMER SUCCESS

Catherine Metzgar, Phd, RD
CLINICAL TEAM

Marlia Braun, PhD, RD
CLINICAL TEAM

Anna Barnwell, MSW, MPH
CLINICAL TEAM

Brent Creighton, PhD
CLINICAL TEAM

We're Hiring

https://www.virtahealth.com/careers

Open positions

Clinical Intake Specialist (Part-Time Contractor)
Denver, CO or Remote

Community Manager
San Francisco

Customer Success Manager, Health Plans
San Francisco

Data Scientist, Machine Learning
San Francisco

Enterprise Partnerships Associate
San Francisco

...

Software Engineer, Backend
San Francisco

Software Engineer, Data
San Francisco

Software Engineer, Full Stack
San Francisco

Nasir Bhanpuri

Improvements in Song Quality with Bombadil Musicians

Quality as judged by musicians and fans

Why isn’t Bombadil playing my favorite songs? Why so many love songs?

Bombadil

- Three-piece folk-pop band
- guitar, bass, piano, and drums
  - Trumpet, accordion, harmonica...
- Rock, ballads, folk, rap (a little)

Nasir Bhanpuri

Variation in Popularity (Spotify streams)

Song	Streams
Thank You	> 1M
Sunny December	> 400 K
A Question	> 300 K
Reasons	> 300 K
Amy’s Friend	> 200 K

BOMBADIL

Nasir Bhanpuri

Machine Learning to Drive Action and Decision-Making

- **Goal:** Predict song popularity
- **Questions:** Which songs will be hits? *Why*? How to improve songs?
- **Users & Collaborators:** Musicians
- **Model Requirements:**
    1. Easy to communicate to musicians
    2. Accuracy
- **Model Benefits:** Ranking, insight into underlying factors, modifications during song development

Nasir Bhanpuri

Feature Selection & Algorithm Selection

Components

- Daniel, James, etc. vocals (1-5)
- Guitar, keyboard, drums, special instruments (1-5)
- Singing, rapping talking (1-5)

Style

- Topic (e.g. love, death…)
- Emotion (e.g. happy, angry...)
- Tempo
- Key & Note

Based on Meaningful Song Characteristics

- 39 Features (Characteristics)
- Feature selection for **Top 20**
    - Lasso regression
- Linear regression (*polynomial terms*)
- $R^2 = 0.6$
    - (Predict popularity in test set)

Nasir Bhanpuri

Model Results: Insights—Feature Directionality*

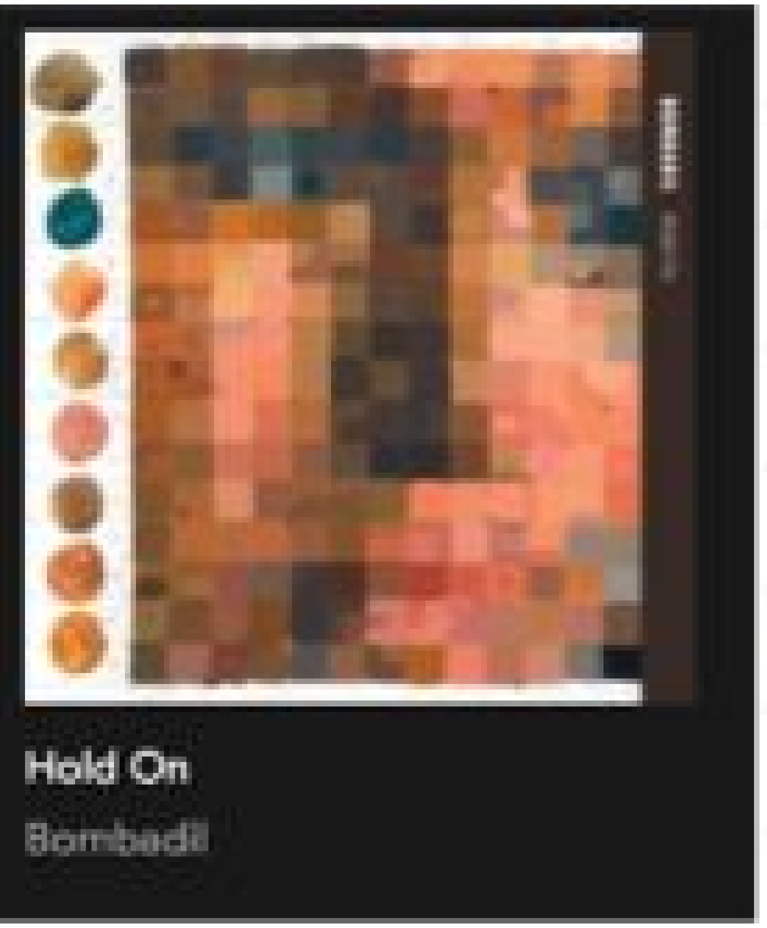
Characteristic	Optimal Value (or Range)
Daniel Vocals	3-4 (Not 5 => tough convo)
James Vocals	≥ 2
Rap and/or Talking	≥ 1
Number of Sections	4
…	

*These are "suggestive" insights from the data and not definitive

More hypothesis generation than conclusive

Experimentation can help confirm

Model Validation Part I: 2015 album *Hold On*



Love Is Simply	451
Seth (Guess I'll Know When I Die)	474
Forgive Me Darling	503
I Can't Believe in Myself and Love You...	622
Honest	675
Rhapsody in Black and White	692
Love You Too Much	727
Amy's Friend	2.6k
Framboise	3.8k
Sunny December	6.1k

- Correctly Rank 4 of top 5
  - Not very good after that
- Good at predicting "hits"
- Still *skepticism*

Actionable Insights

Characteristic	Optimal Value (or Range)	Translate to songwriting
Daniel Vocals	3-4 (Not 5 => tough convo)	More mixed vocals (not all one)
James Vocals	≥ 2	Bridge section helps
Rap and/or Talking	≥ 1	...
Number of Sections	4	
...		

Model Validation Part II: A/B Test

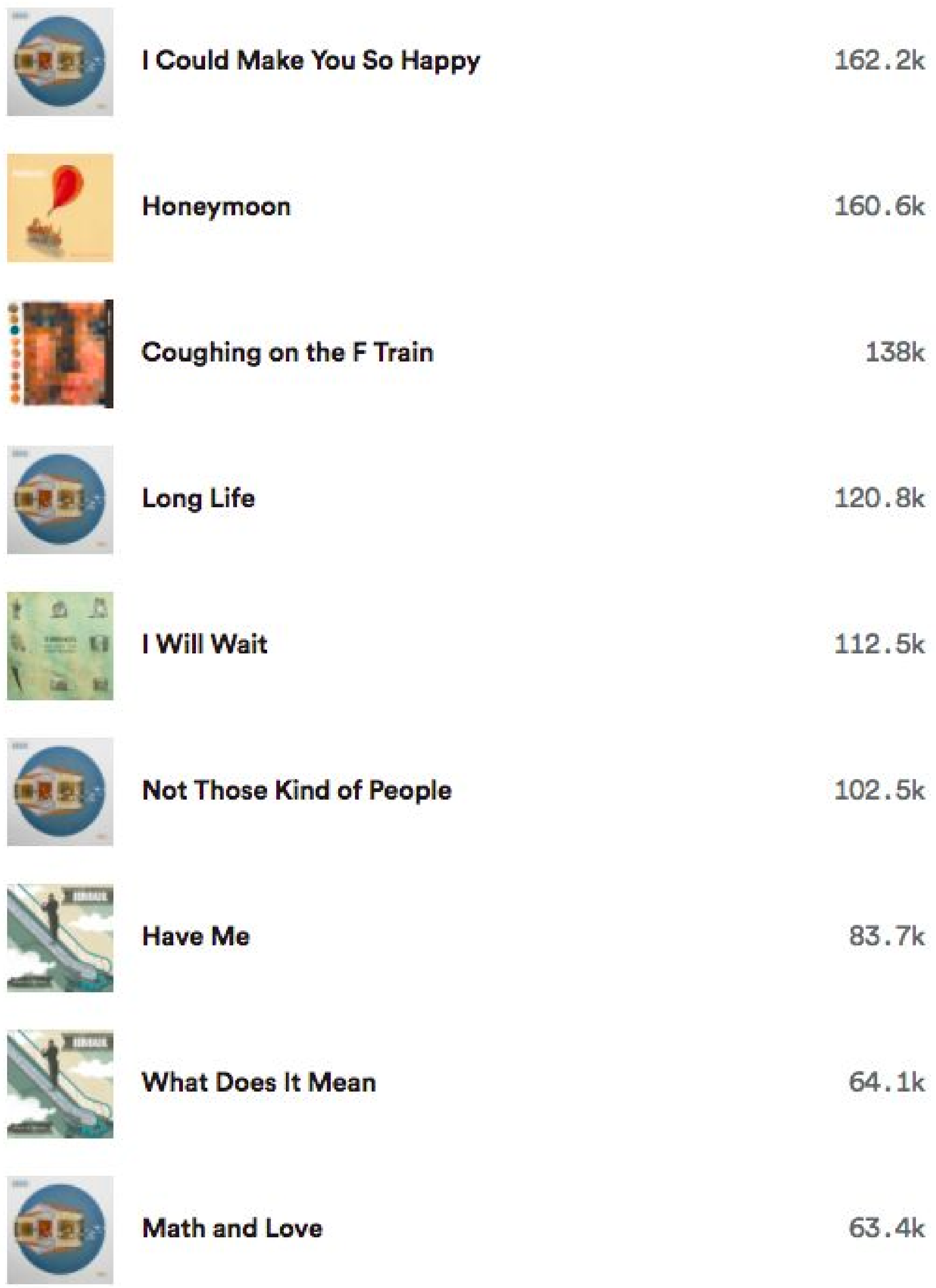
A/B Test to see if it works!

Let’s try?!

- Randomized order of 2 versions
- > 1000 responses
- **59% prefer new song**

Impact

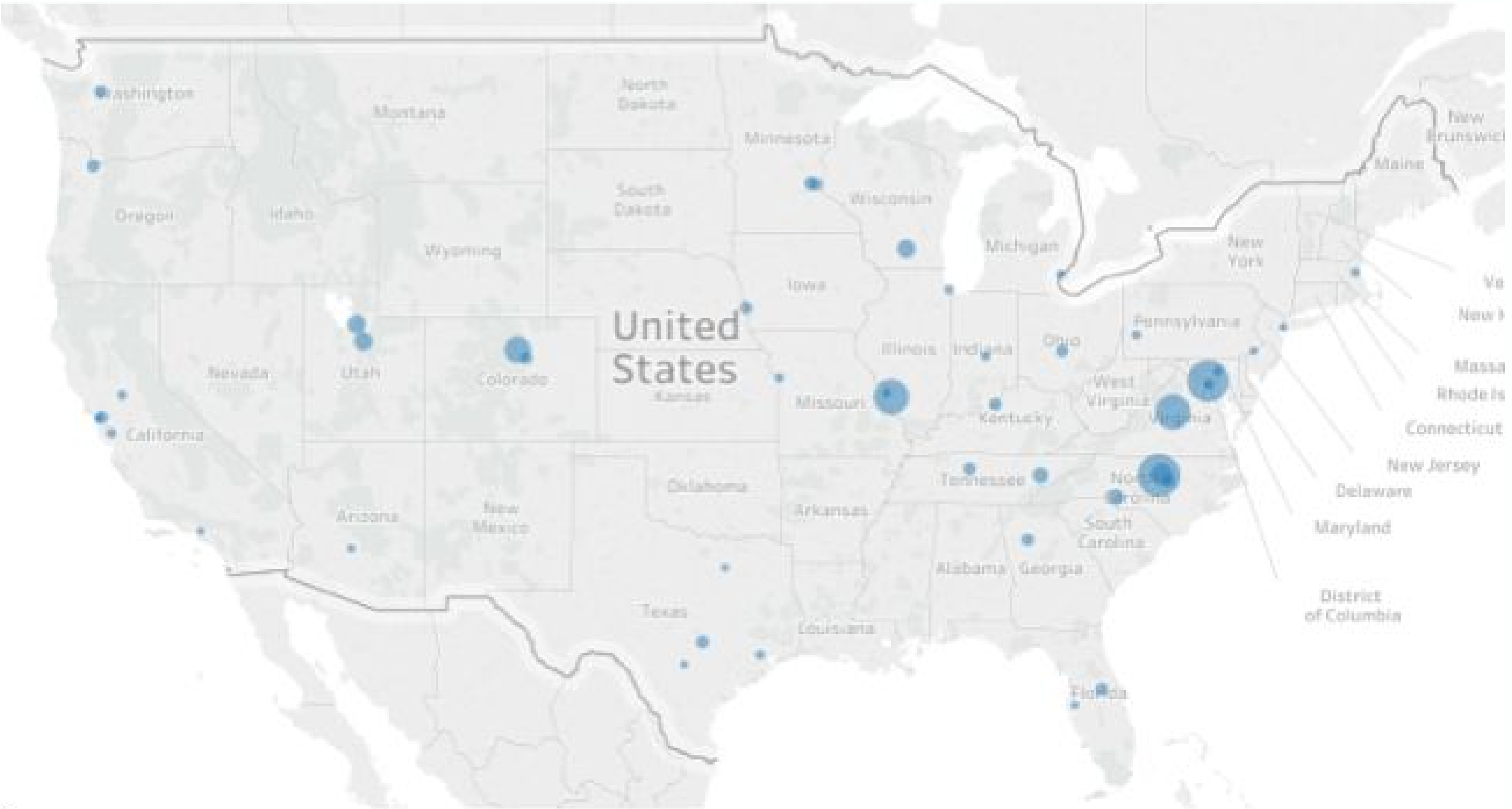
- Disrupted song-writing (in a good way!)
- "I Could Make you so Happy"
  - Band, fans, strangers prefer data version
- Correctly predicted "Hits"
  - Top 3 of 4 songs on latest album
- Successfully added "Perfect" to album
- *Important caveats*
  - [Expectedly] Very wrong prediction with sparse data (human + machine > machine)
  - Rely on band creativity
  - Promotions



- Rank correlation
  - P<0.05, rho = 0.67
  - (Ignoring Long Life)

Now what?

- Other analytics
  - Concert destinations → Successful Denver trip!
- Demographics to target
  - Most popular among college-aged



Thanks Bombadil!

James Phillips - *percussion, vocals* Daniel Michalak - *piano, guitar, vocals* Stacy Harden - *upright bass, vocals*

with Bryan Rahija - *guitar*

Additional players: Andrew Maguire - *percussion* John Vanderslice - *minimoog* Nasir Bhanpuri - *data science*

ML with humans & for humans when…

- Model development & iteration benefit from human expertise
  - Data are limited and/or accelerate development
- Insights & interpretability are valuable
  - When more important than accuracy, favor “transparent” models
- Output inform decisions
  - Human + machine > machine (at least sometimes!)

Thank you!

Arivoli (Oli) Tirouvingadame

Jayaradha Natarajan

Data Riders

Qventus

NFL.com, Virta Health, Bombadil, fivethirtyeight.com

We’re Hiring

https://www.virtahealth.com/careers

Open positions

Clinical Intake Specialist (Part-Time Contractor)
Denver, CO or Remote

Community Manager
San Francisco

Customer Success Manager, Health Plans
San Francisco

Data Scientist, Machine Learning
San Francisco

Enterprise Partnerships Associate
San Francisco

…

Software Engineer, Backend
San Francisco

Software Engineer, Data
San Francisco

Software Engineer, Full Stack
San Francisco