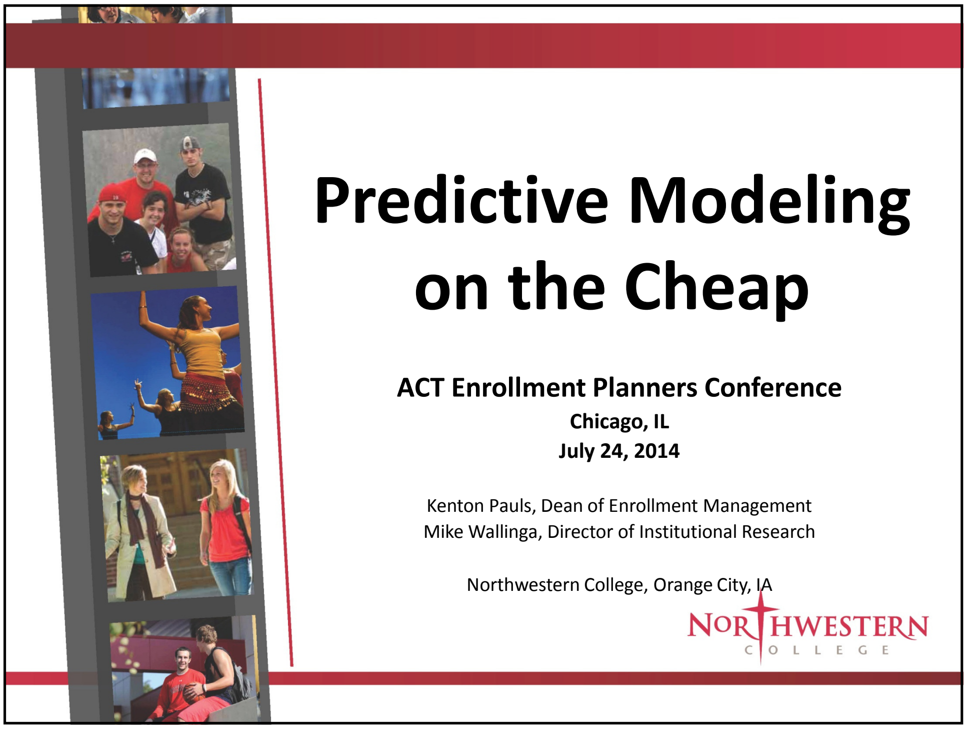


Predictive Modeling on the Cheap

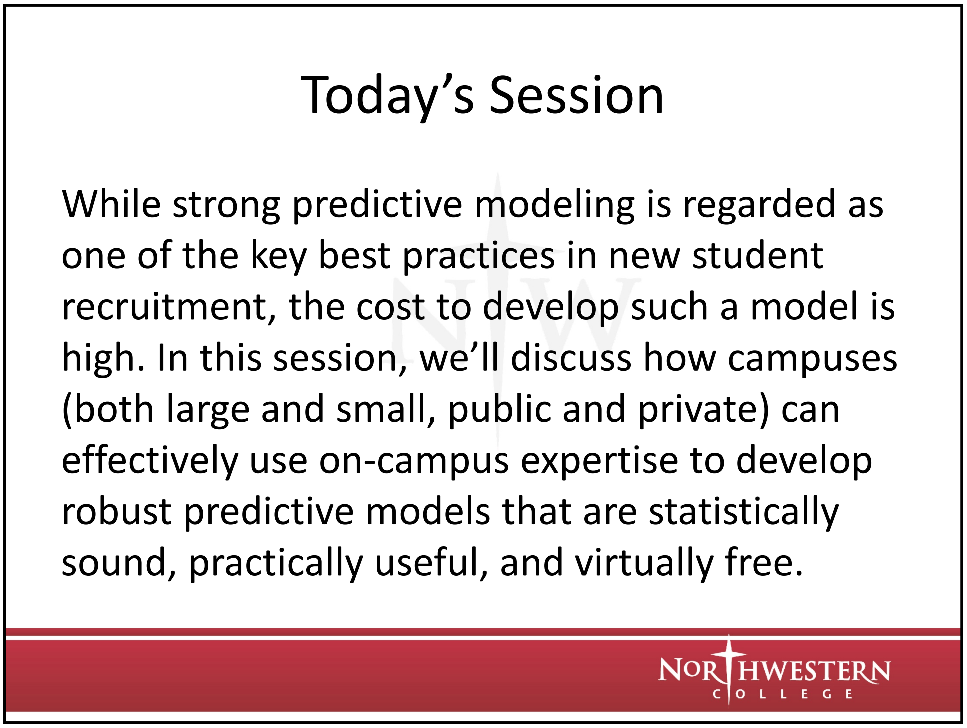
ACT Enrollment Planners Conference

Chicago, IL

July 24, 2014

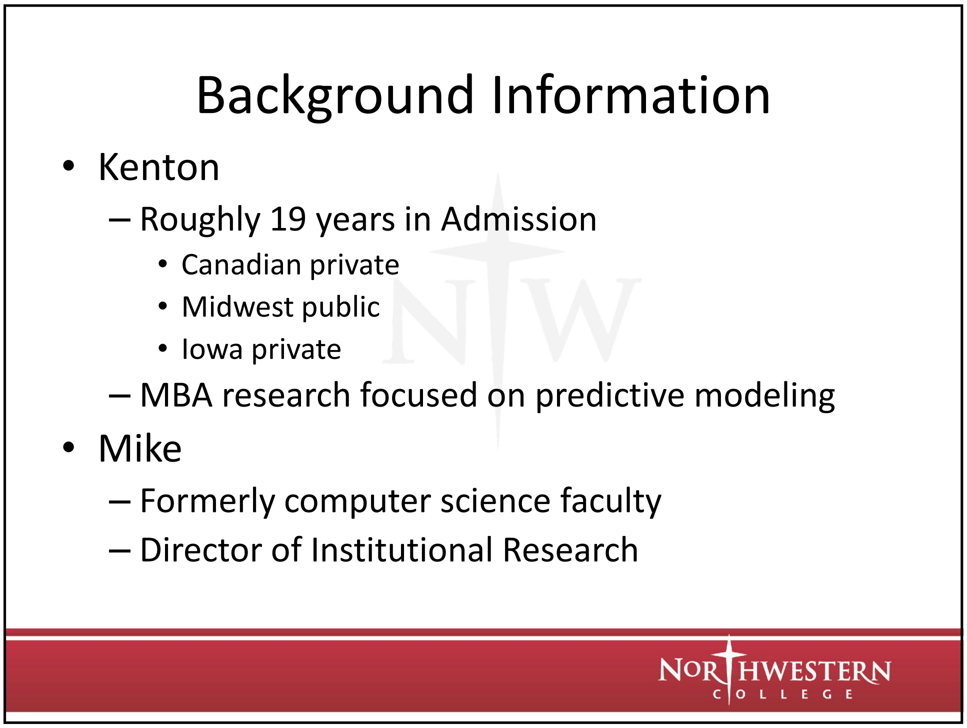
Kenton Pauls, Dean of Enrollment Management
Mike Wallinga, Director of Institutional Research

Northwestern College, Orange City, IA



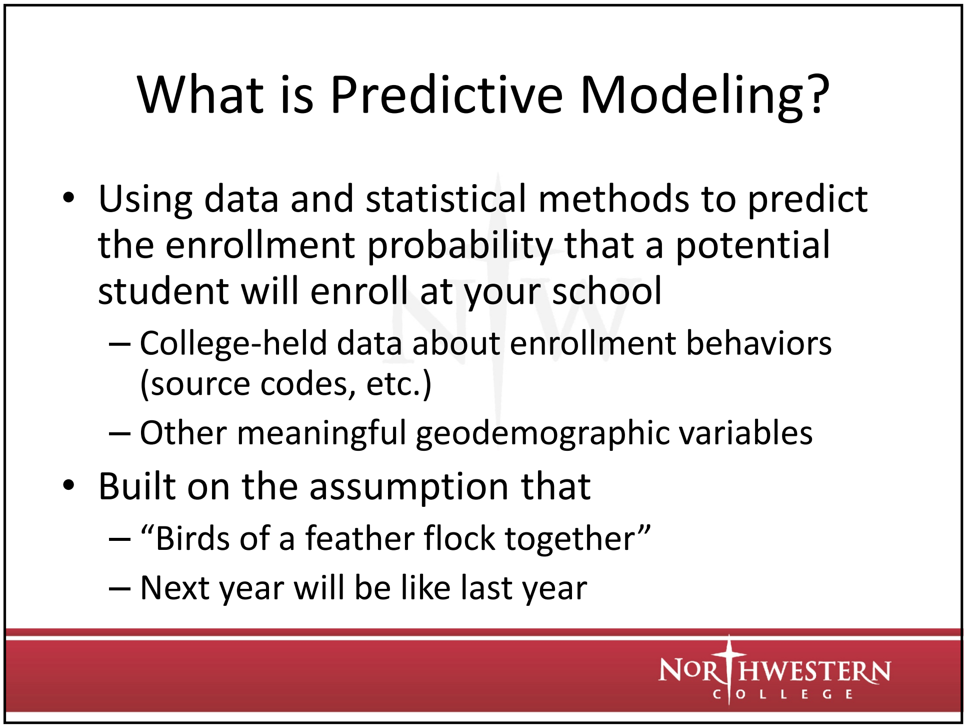
Today’s Session

While strong predictive modeling is regarded as one of the key best practices in new student recruitment, the cost to develop such a model is high. In this session, we’ll discuss how campuses (both large and small, public and private) can effectively use on-campus expertise to develop robust predictive models that are statistically sound, practically useful, and virtually free.



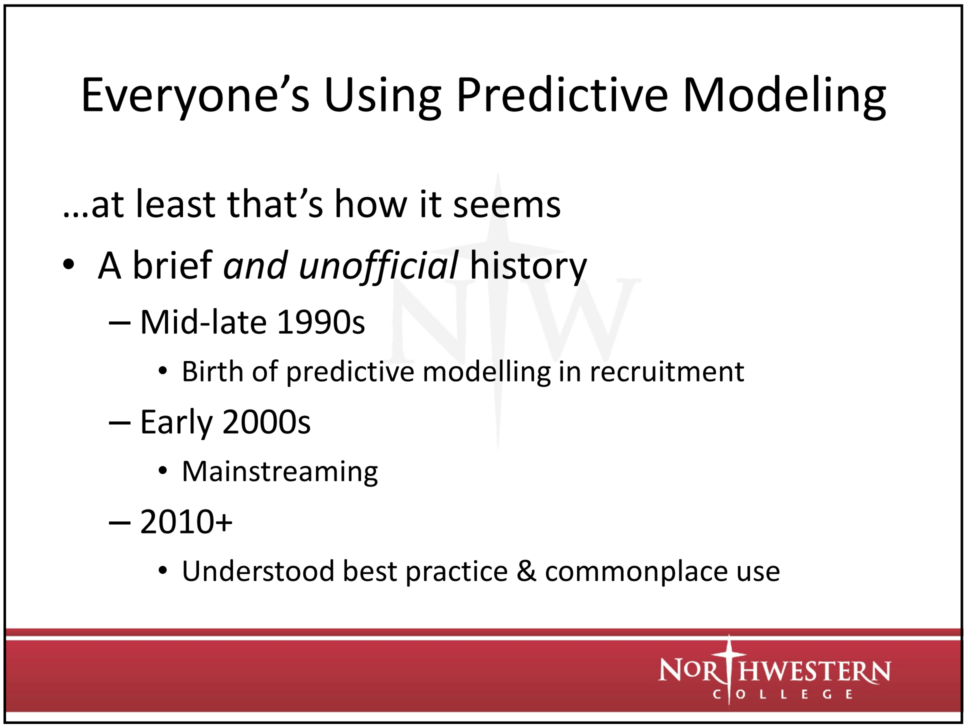
Background Information

- Kenton
  - Roughly 19 years in Admission
    - Canadian private
    - Midwest public
    - Iowa private
  - MBA research focused on predictive modeling
- Mike
  - Formerly computer science faculty
  - Director of Institutional Research



What is Predictive Modeling?

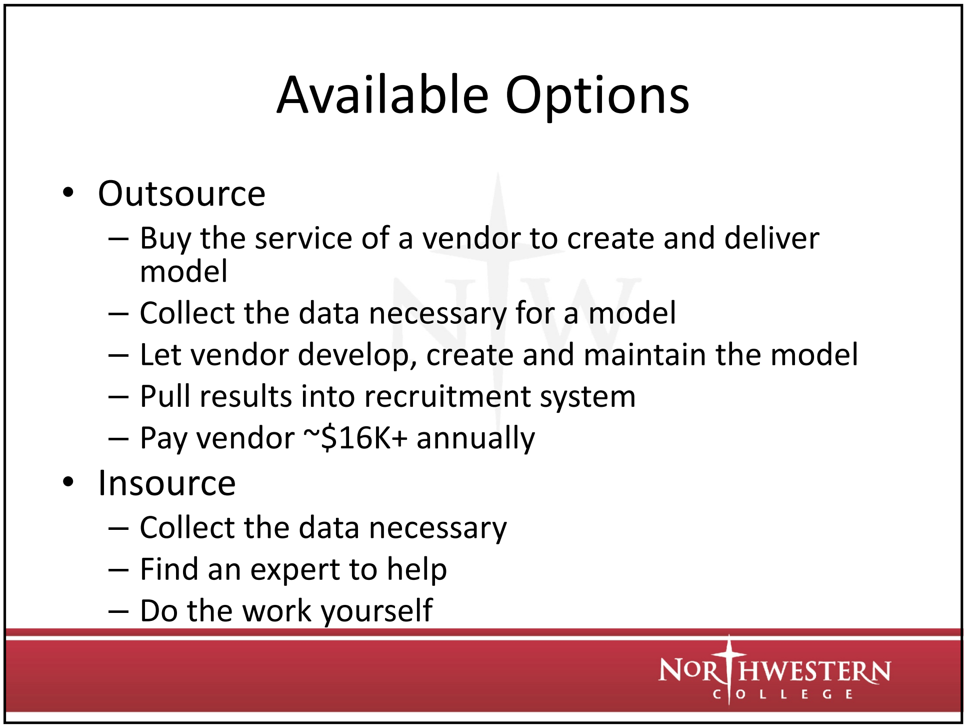
- Using data and statistical methods to predict the enrollment probability that a potential student will enroll at your school
  - College-held data about enrollment behaviors (source codes, etc.)
  - Other meaningful geodemographic variables
- Built on the assumption that
  - “Birds of a feather flock together”
  - Next year will be like last year



Everyone's Using Predictive Modeling

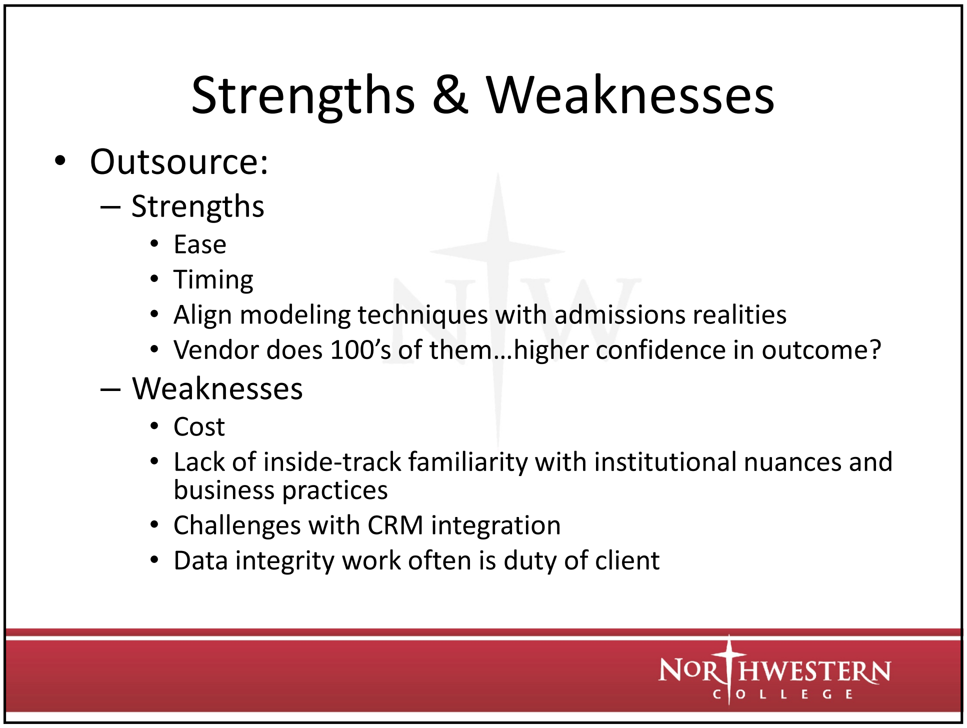
…at least that's how it seems

- A brief *and unofficial* history
  - – Mid-late 1990s
    - Birth of predictive modelling in recruitment
  - – Early 2000s
    - Mainstreaming
  - – 2010+
    - Understood best practice & commonplace use



Available Options

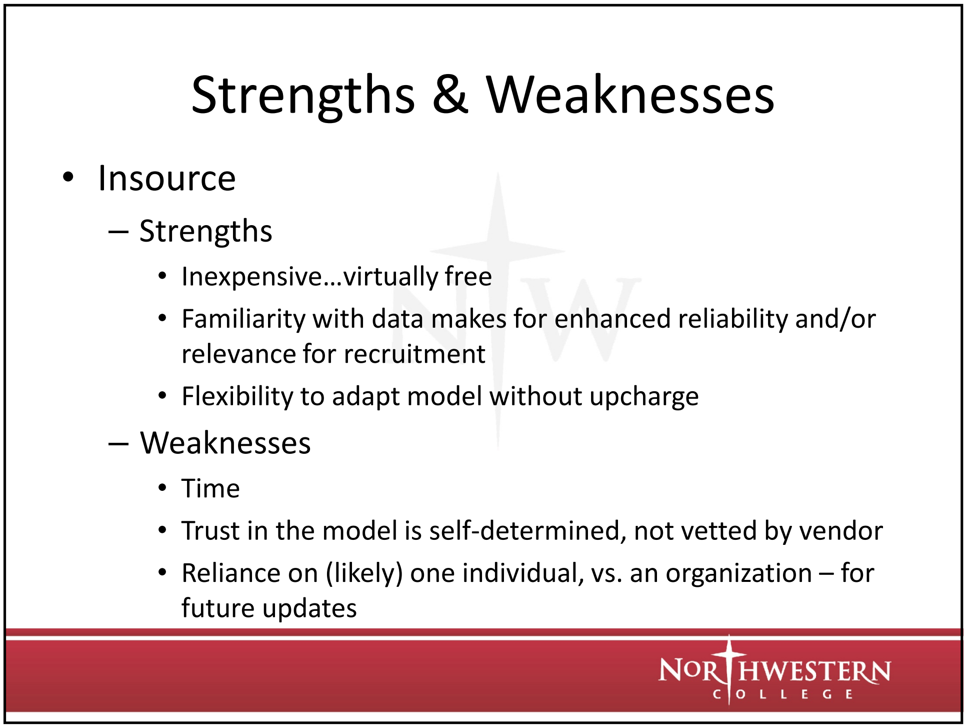
- Outsource
  - – Buy the service of a vendor to create and deliver model
  - – Collect the data necessary for a model
  - – Let vendor develop, create and maintain the model
  - – Pull results into recruitment system
  - – Pay vendor ~$16K+ annually
- Insource
  - – Collect the data necessary
  - – Find an expert to help
  - – Do the work yourself



Strengths & Weaknesses

- Outsource:
  - Strengths
    - Ease
    - Timing
    - Align modeling techniques with admissions realities
    - Vendor does 100's of them…higher confidence in outcome?
  - Weaknesses
    - Cost
    - Lack of inside-track familiarity with institutional nuances and business practices
    - Challenges with CRM integration
    - Data integrity work often is duty of client

NORTHWESTERN COLLEGE



Strengths & Weaknesses

- Insource
  - Strengths
    - Inexpensive…virtually free
    - Familiarity with data makes for enhanced reliability and/or relevance for recruitment
    - Flexibility to adapt model without upcharge
  - Weaknesses
    - Time
    - Trust in the model is self-determined, not vetted by vendor
    - Reliance on (likely) one individual, vs. an organization – for future updates

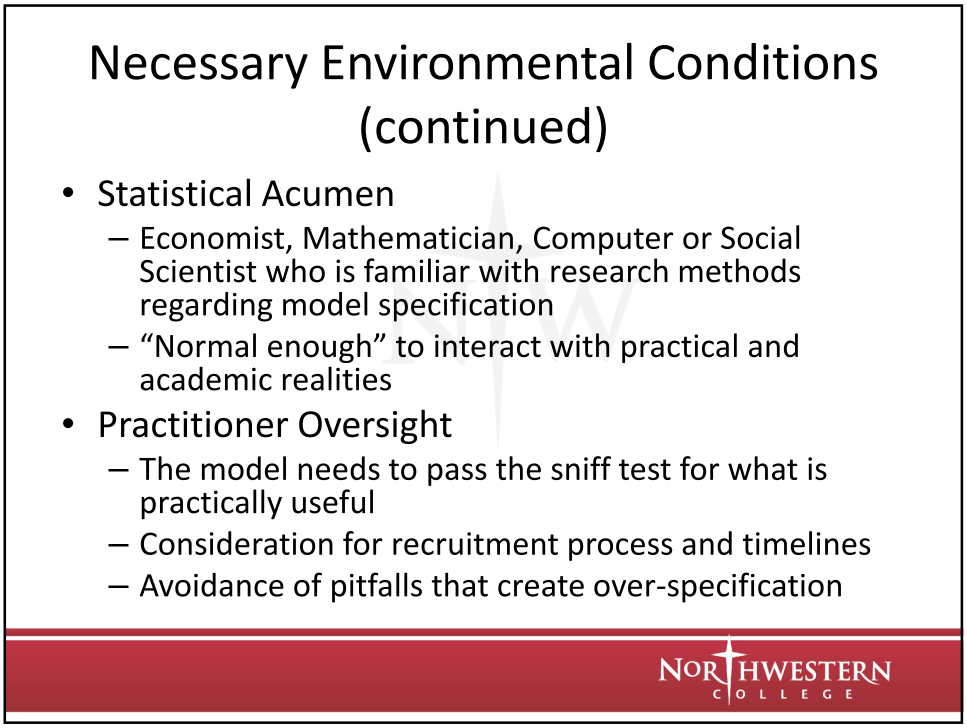
NORTHWESTERN COLLEGE

Big Data @ Small College?

- It is possible for any college
  - Large & small

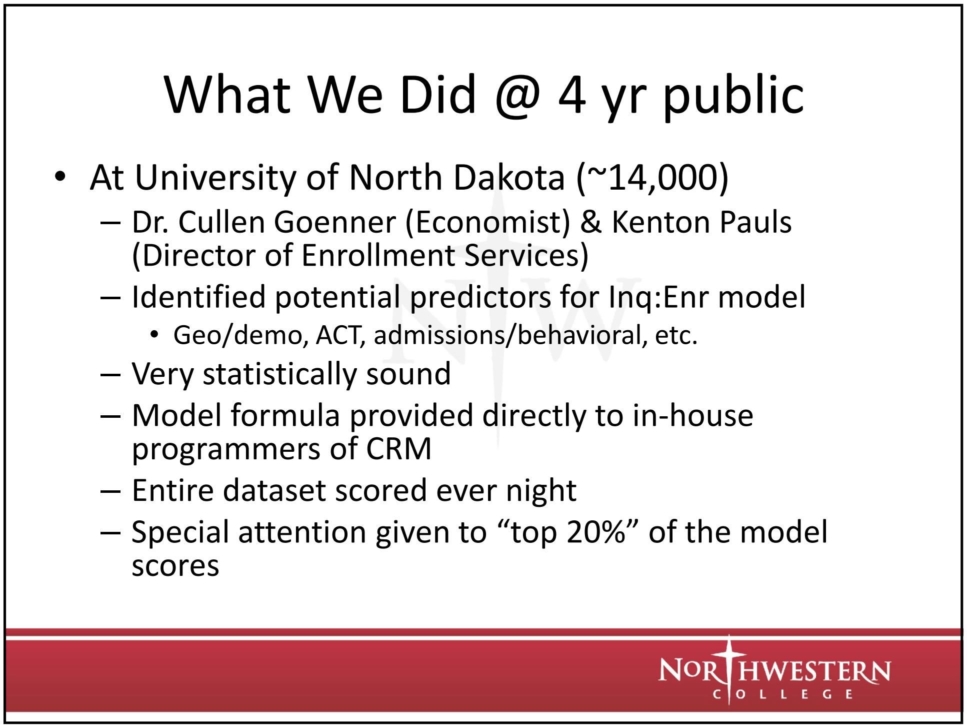
Necessary Environmental Conditions for Success

- Clean and old data is a must
  - Two years+ minimally
  - Behavioral data (source codes)
    - One or many sources
    - Date-sensitive source codes, etc.
  - Funnel history
  - Culture of data capture
- Technical baseline
  - Turn potentially meaningful data into 1’s and 0’s
  - Willingness to provide these data repeatedly and much more data than will be used (for research purposes)



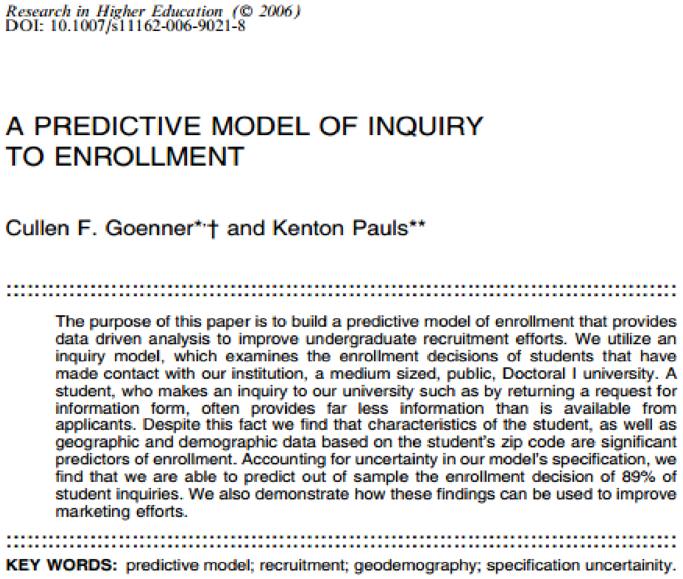
Necessary Environmental Conditions (continued)

- Statistical Acumen
  - Economist, Mathematician, Computer or Social Scientist who is familiar with research methods regarding model specification
  - "Normal enough" to interact with practical and academic realities
- Practitioner Oversight
  - The model needs to pass the sniff test for what is practically useful
  - Consideration for recruitment process and timelines
  - Avoidance of pitfalls that create over-specification



What We Did @ 4 yr public

- At University of North Dakota (~14,000)
  - Dr. Cullen Goenner (Economist) & Kenton Pauls (Director of Enrollment Services)
  - Identified potential predictors for Inq:Enr model
    - Geo/demo, ACT, admissions/behavioral, etc.
  - Very statistically sound
  - Model formula provided directly to in-house programmers of CRM
  - Entire dataset scored ever night
  - Special attention given to "top 20%" of the model scores



Research in Higher Education (© 2006)
DOI: 10.1007/s11162-006-9021-8

A PREDICTIVE MODEL OF INQUIRY TO ENROLLMENT

Cullen F. Goenner*,† and Kenton Pauls**

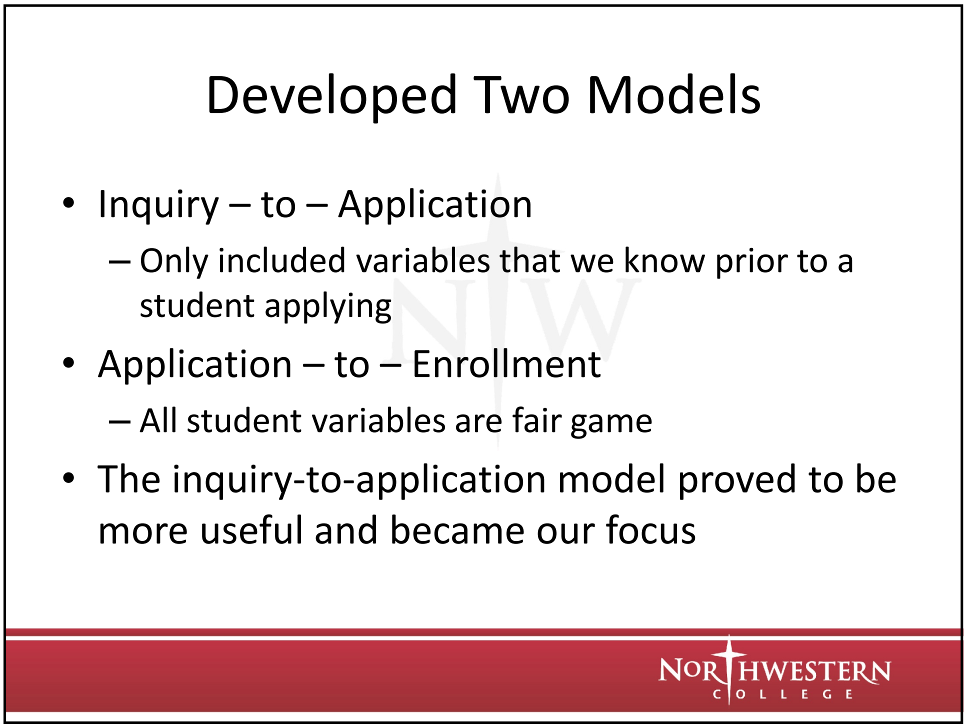
The purpose of this paper is to build a predictive model of enrollment that provides data driven analysis to improve undergraduate recruitment efforts. We utilize an inquiry model, which examines the enrollment decisions of students that have made contact with our institution, a medium sized, public, Doctoral I university. A student, who makes an inquiry to our university such as by returning a request for information form, often provides far less information than is available from applicants. Despite this fact we find that characteristics of the student, as well as geographic and demographic data based on the student's zip code are significant predictors of enrollment. Accounting for uncertainty in our model's specification, we find that we are able to predict out of sample the enrollment decision of 89% of student inquiries. We also demonstrate how these findings can be used to improve marketing efforts.

KEY WORDS: predictive model; recruitment; geodemography; specification uncertainity.

http://goo.gl/5NHhcF

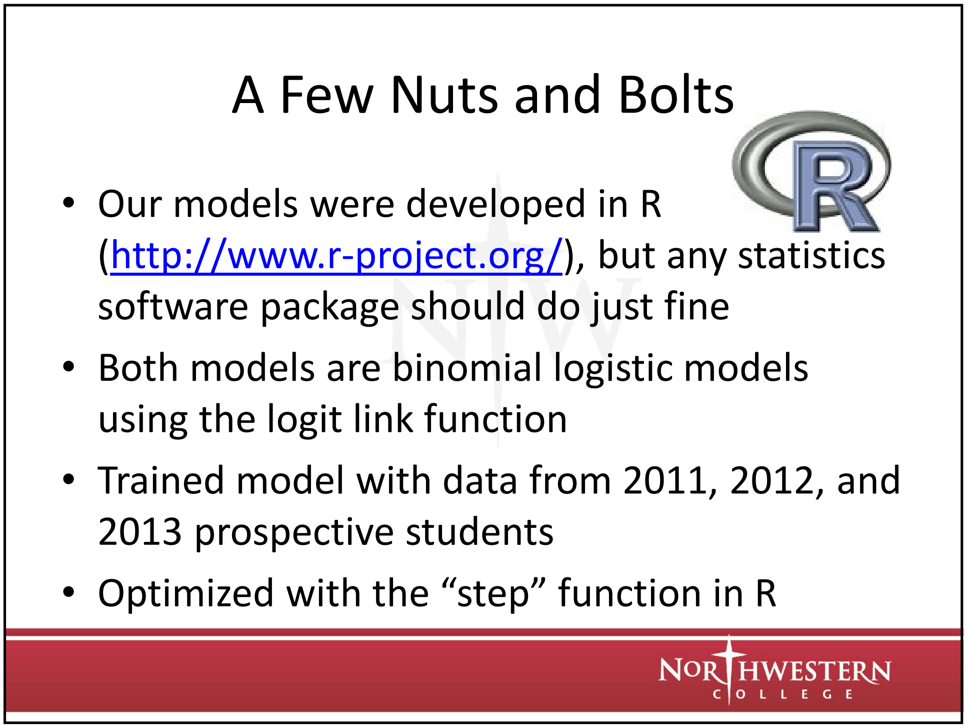
What We Did @ 4 yr private

- At Northwestern College, IA (~1,200)
  - Mike Wallinga (Director of IR, CS/Math background)
  - Very familiar with institutional data
  - Savvy in both enrollment mgt & academic matters
  - Fixed data capture mechanisms (~2 yrs)
  - Developed model
  - Mike scores the model weekly
  - Full integration into CRM with special emphasis on top third



Developed Two Models

- Inquiry – to – Application
  - Only included variables that we know prior to a student applying
- Application – to – Enrollment
  - All student variables are fair game
- The inquiry-to-application model proved to be more useful and became our focus

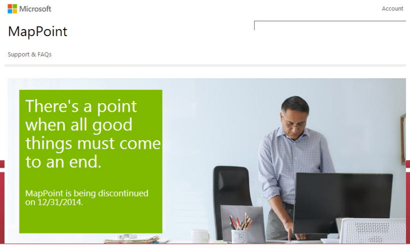


A Few Nuts and Bolts

- Our models were developed in R (http://www.r-project.org/), but any statistics software package should do just fine
- Both models are binomial logistic models using the logit link function
- Trained model with data from 2011, 2012, and 2013 prospective students
- Optimized with the “step” function in R

Data Preparation

- Collected basic data about students from our administrative database
- Supplemented with zip code-level demographics
  - Taken from Microsoft MapPoint, but could get it from US Census Bureau and other sources, too

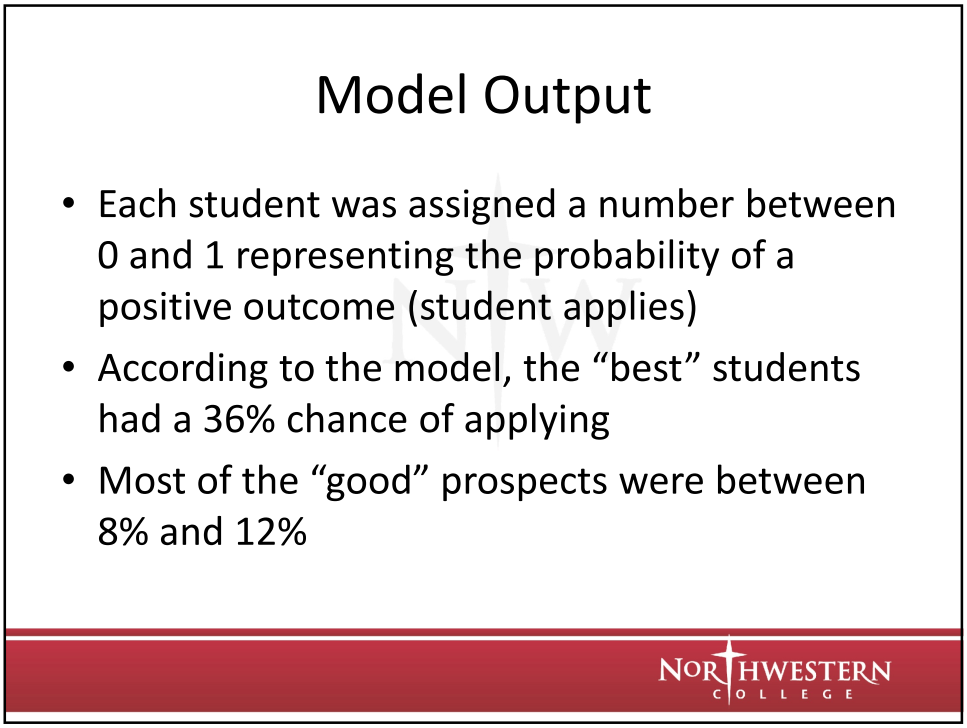


There's a point when all good things must come to an end.

MapPoint is being discontinued on 12/31/2014.

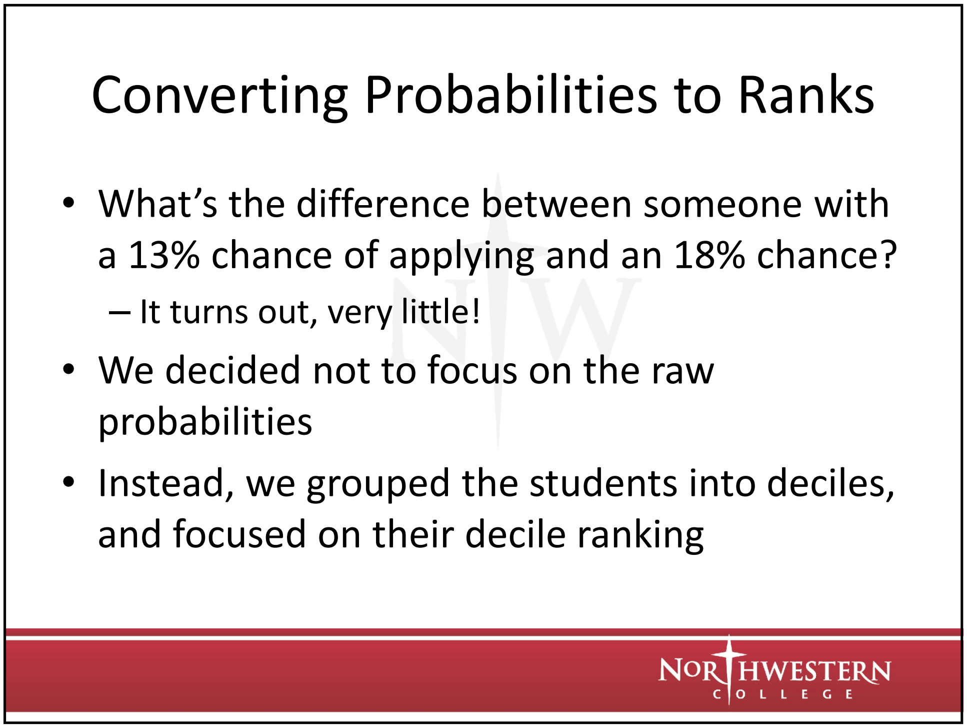
Included Variables

- Days since initial inquiry status *
- First contact method
- ~~Gender~~
- Home state ("Common states") *
- Campus visit
- ~~Population density~~
- ~~Average household income~~
- Average household expenditure on education
- Average household expenditure on reading *
- ~~Median housing value~~
- % of adults with some college *
- ~~% of adults with bachelor's~~
- % of adults with grad degree
- % of population, white ethnicity
- ~~% of population, age 15-24~~
- ~~Average annual temperature difference between home zip code and Orange City~~



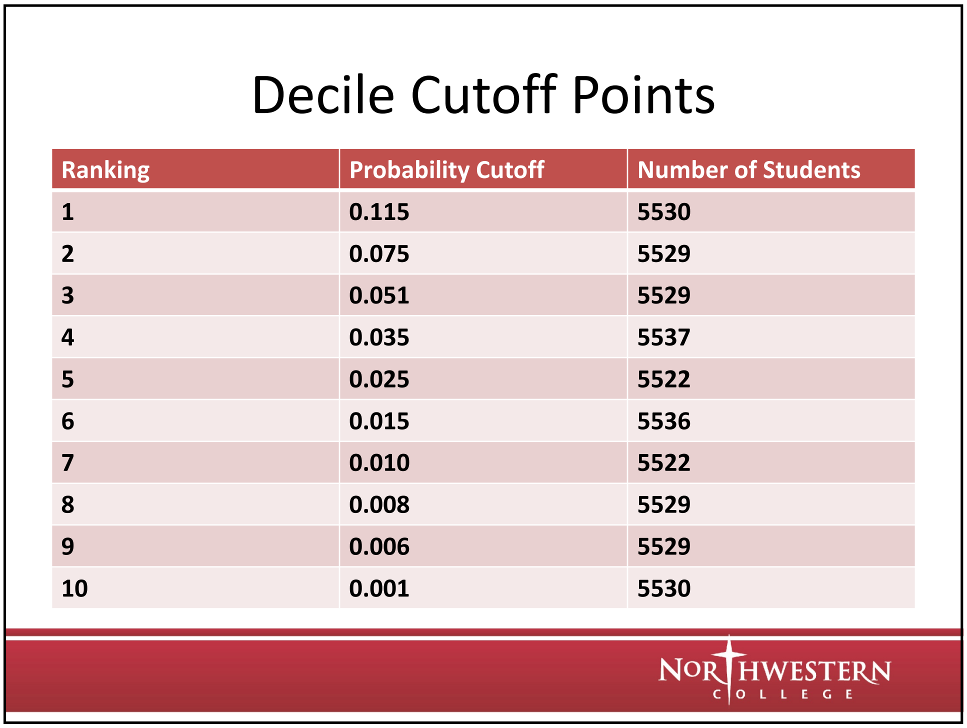
Model Output

- Each student was assigned a number between 0 and 1 representing the probability of a positive outcome (student applies)
- According to the model, the “best” students had a 36% chance of applying
- Most of the “good” prospects were between 8% and 12%



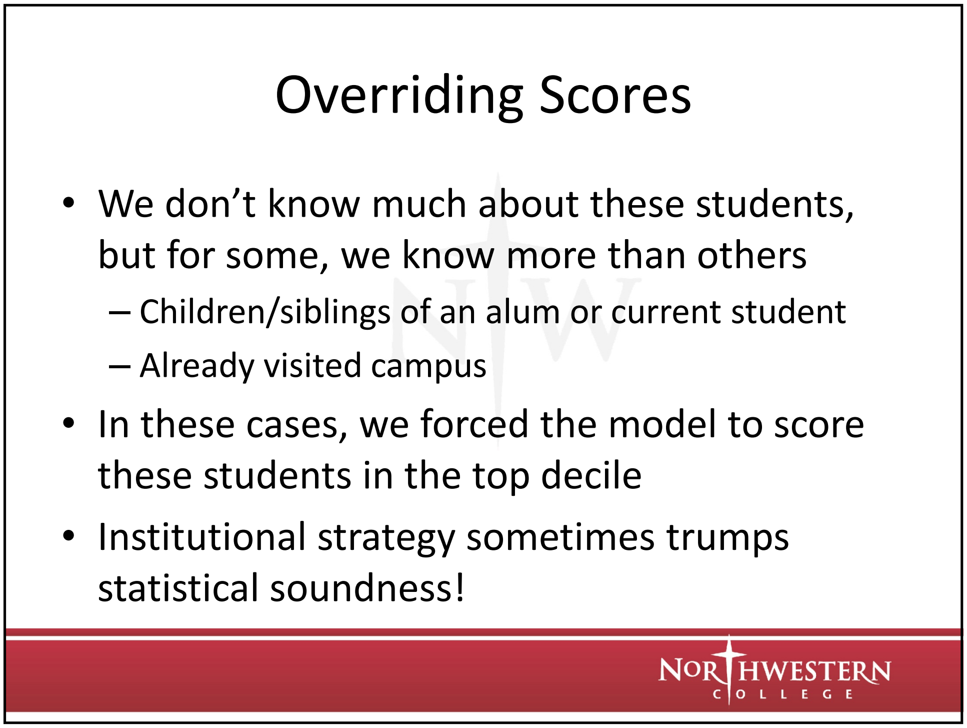
Converting Probabilities to Ranks

- What’s the difference between someone with a 13% chance of applying and an 18% chance?
  - It turns out, very little!
- We decided not to focus on the raw probabilities
- Instead, we grouped the students into deciles, and focused on their decile ranking



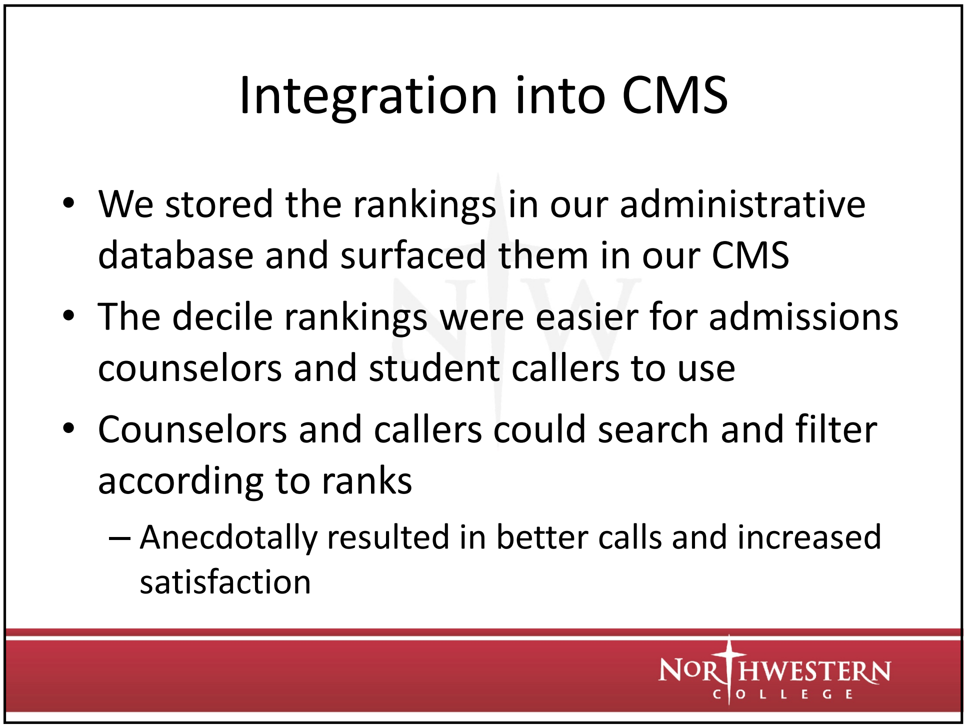
Decile Cutoff Points

Ranking	Probability Cutoff	Number of Students
1	0.115	5530
2	0.075	5529
3	0.051	5529
4	0.035	5537
5	0.025	5522
6	0.015	5536
7	0.010	5522
8	0.008	5529
9	0.006	5529
10	0.001	5530



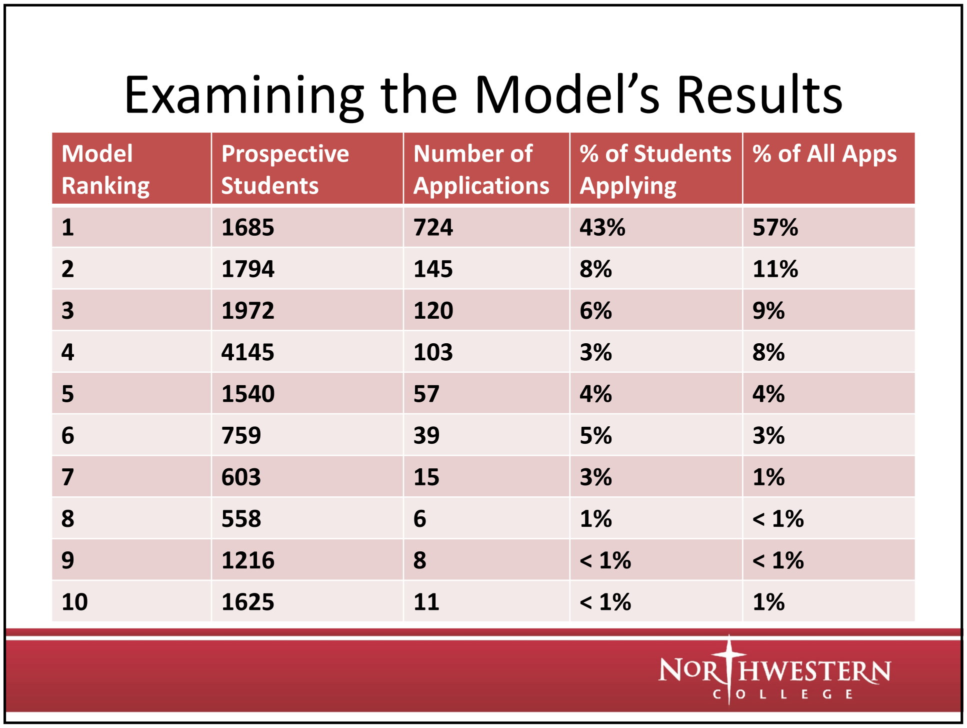
Overriding Scores

- We don’t know much about these students, but for some, we know more than others
  - Children/siblings of an alum or current student
  - Already visited campus
- In these cases, we forced the model to score these students in the top decile
- Institutional strategy sometimes trumps statistical soundness!



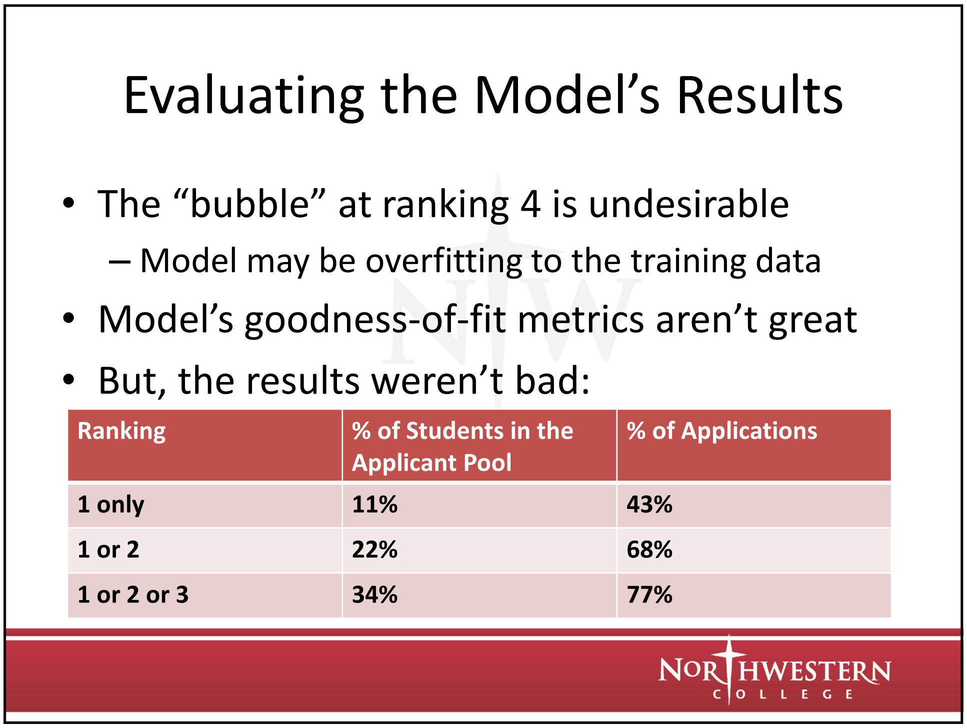
Integration into CMS

- We stored the rankings in our administrative database and surfaced them in our CMS
- The decile rankings were easier for admissions counselors and student callers to use
- Counselors and callers could search and filter according to ranks
  - Anecdotally resulted in better calls and increased satisfaction



Examining the Model's Results

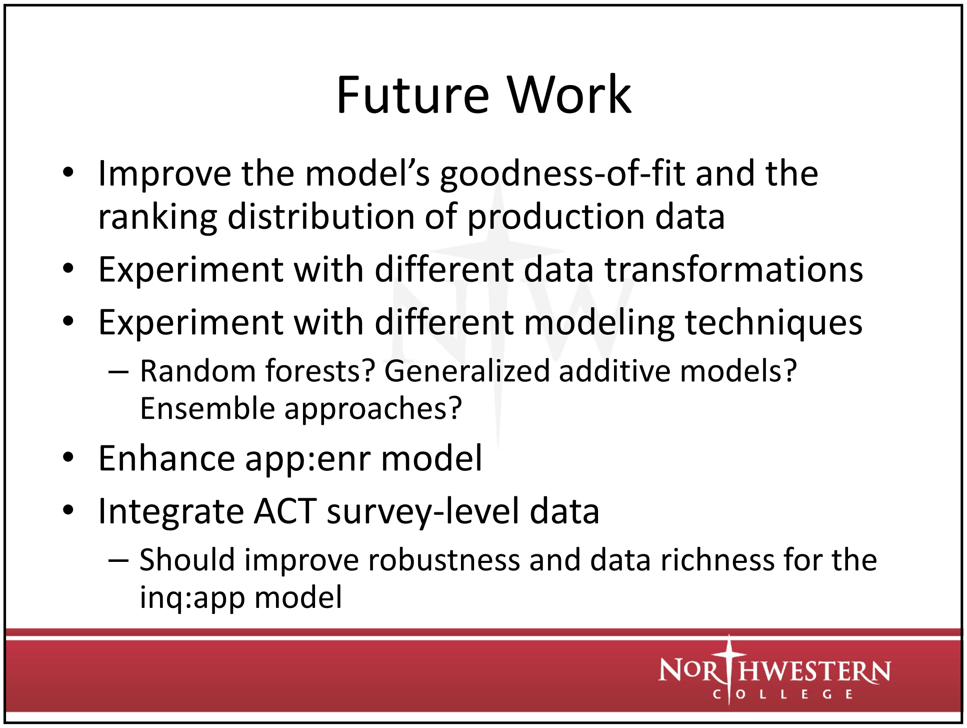
Model Ranking	Prospective Students	Number of Applications	% of Students Applying	% of All Apps
1	1685	724	43%	57%
2	1794	145	8%	11%
3	1972	120	6%	9%
4	4145	103	3%	8%
5	1540	57	4%	4%
6	759	39	5%	3%
7	603	15	3%	1%
8	558	6	1%	< 1%
9	1216	8	< 1%	< 1%
10	1625	11	< 1%	1%



Evaluating the Model’s Results

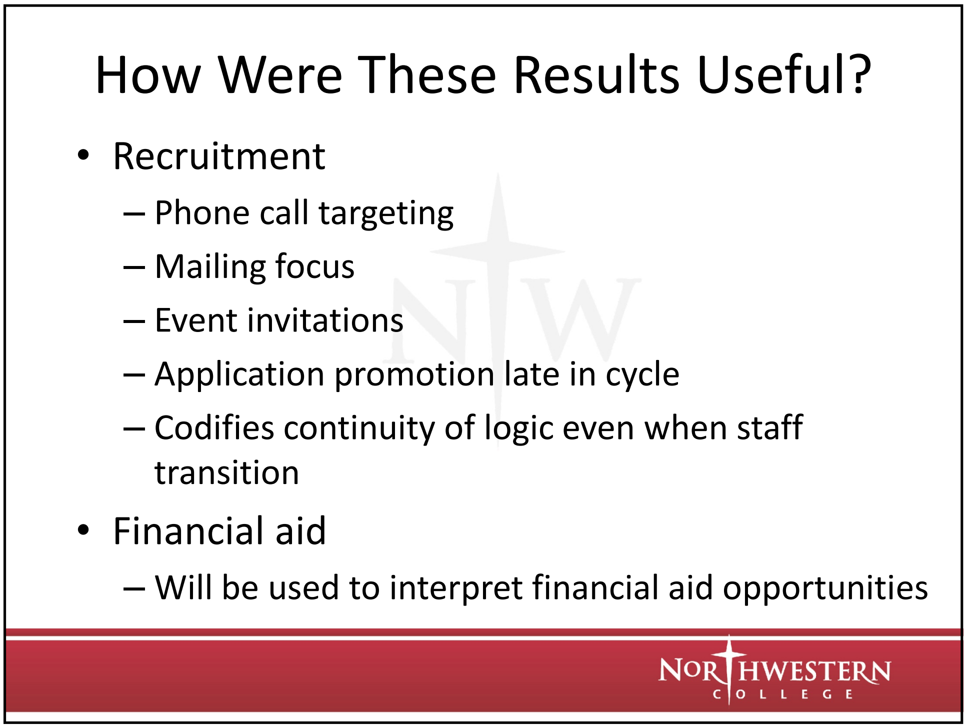
- The “bubble” at ranking 4 is undesirable
  - Model may be overfitting to the training data
- Model’s goodness-of-fit metrics aren’t great
- But, the results weren’t bad:

Ranking	% of Students in the Applicant Pool	% of Applications
1 only	**11%**	**43%**
1 or 2	**22%**	**68%**
1 or 2 or 3	**34%**	**77%**



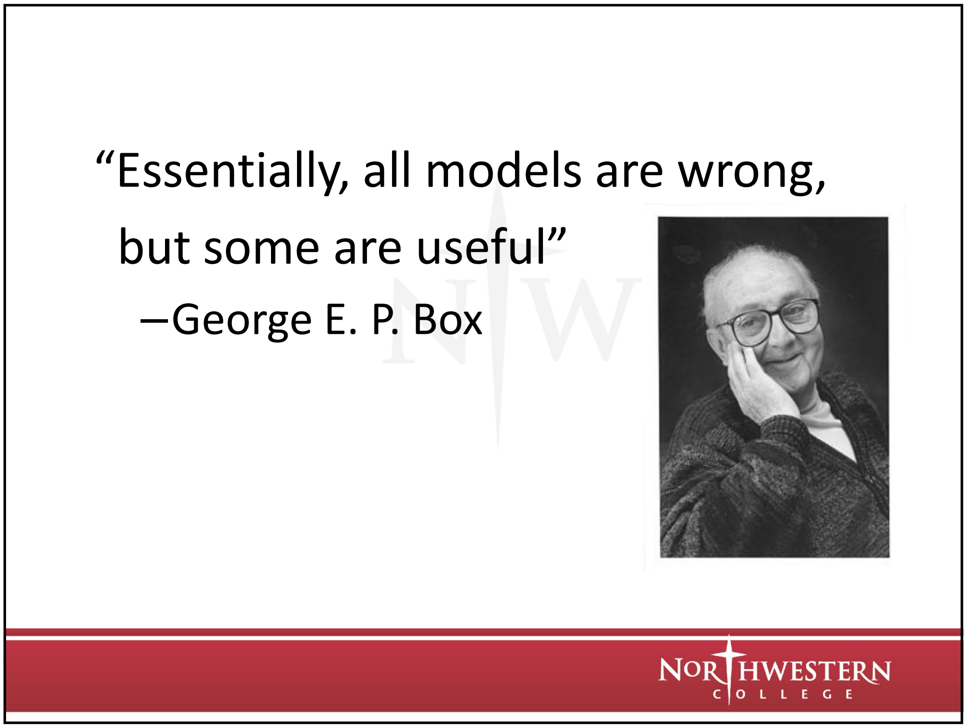
Future Work

- Improve the model’s goodness-of-fit and the ranking distribution of production data
- Experiment with different data transformations
- Experiment with different modeling techniques
  - Random forests? Generalized additive models? Ensemble approaches?
- Enhance app:enr model
- Integrate ACT survey-level data
  - Should improve robustness and data richness for the inq:app model



How Were These Results Useful?

- Recruitment
  - Phone call targeting
  - Mailing focus
  - Event invitations
  - Application promotion late in cycle
  - Codifies continuity of logic even when staff transition
- Financial aid
  - Will be used to interpret financial aid opportunities



“Essentially, all models are wrong, but some are useful”

–George E. P. Box

Questions & Discussion?

Our contact info

kenton.pauls@nwciowa.edu

mwalling@nwciowa.edu