PREDICTION MODELING FOR ACADEMIC SUCCESS

Scott L. Bruce, EdD, ATC
Associate Professor/Co-clinical Coordinator/Director of Research
Masters of Athletic Training Program
Arkansas State University

2019 ASAHP Annual Conference – October 2019

Conflict of Interest

I have no affiliation or financial interest with or involvement in any organization or entity that will be discussed as part of this presentation.

Objectives

- To identify factors which may predict academic success
- To evaluate factors to decide which may contribute to predicting academic success
- To create a prediction model that fits the learners' specific situation
- To evaluate whether or not their prediction model has the ability to identify the most qualified candidates

“Begin with the end in mind”
(Covey, 2004)

- What is the ultimate indicator of student success in an academic program?
  - In Athletic Training it is 1st-attempt Board of Certification (BOC) exam success
- Second question: How does one go about predicting who might be successful in a graduate academic program?
  - Prediction modeling
- For today’s presentation, I will use a Professional Master’s of Athletic Training Program as my example

Outcome Measures

- Based on:
  - Sensitivity
  - Specificity
  - Odds Ratio
  - Relative Risk???
  - Likelihood Ratios

Odds Ratio

- The odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure
- If the odds are the same for both groups, OR = 1.0

Odds Ratio

- Odds and horse racing
  - 2:1 horse or 50:1 horse?
    - 2:1 is better than the long shot of 50:1
  - 2:1 injury/academic success or 50:1 injury/ academic success
    - 50:1 says you are more likely to get injured or have academic success than someone at 2:1

Relative Risk

- The likelihood that someone who has been exposed to a risk factor will develop the injury as compared to someone who does not have the risk factor
  - If the probability is the same for both groups, RR = 1.0

Relative Frequency of Success

- Relative Frequency of Success (RFS) replaced Relative Risk since risk is not an appropriate term when measuring success

- Relative Frequency of Success is defined as:
  - The likelihood that someone who has the predictor is forecast to be successful in a graduate academic program is successful compared with one who has not been so classified

Likelihood Ratios

- Positive LR (or +LR) is the probability that a student with the predictor (or possesses the predictor), would be successful in a graduate academic program compared to the probability that a student without the variable (or does not possess the predictor) would be successful in a graduate program
- Negative LR (or –LR) is the probability that a student w/o the predictor would be successful in a graduate academic program compared to the probability of the student with the predictor would be successful in a graduate program

https://www.cebm.net/2014/02/likelihood-ratios/

Interpreting Different Statistics

Association	Small	Moderate	Large	Very Large
Odds Ratio	≥ 1.5	≥ 3.4	≥ 9.0	≥ 32.0
Relative Risk/Relative Frequency of Success	≥ 1.1	≥ 1.4	≥ 2.0	≥ 3.3
+ Likelihood Ratio	≤ 5.0	≤ 10.0	> 10.0	
– Likelihood Ratio	≤ 0.5	≤ 0.2	< 0.1	
Hazards Ratio	≥ 1.3	≥ 2.0	≥ 4.0	≥ 10.0

Hopkins WG. In: Verhagen E, van Mechelen W (eds.). Sports Injury Research. 2010.
Jaeschke et al., JAMA, 271(9); 1994

Steps in Prediction Modeling

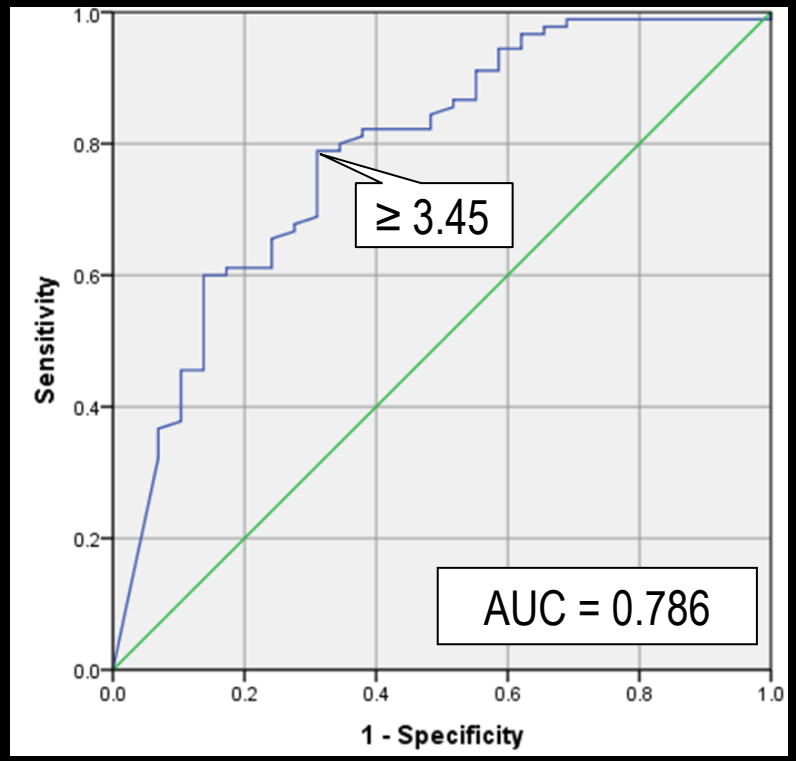
- Three step process
  1. Create the prediction model
     - Clear operational definition of the dependent variable
     - ID any and all potential predictor variables
  2. Determine validity
     - Apply rule to a different population
  3. Conduct impact analysis
     - Evidence rule change behavior, changes outcomes, or reduces costs

Today's focus will be on the development of CPGs

Clear Operational Definition of the Dependent Variable

Association	Small	Moderate	Large	Very Large
Odds Ratio	≥ 1.5	≥ 3.4	≥ 9.0	≥ 32.0
Relative Risk/Relative Frequency of Success	≥ 1.1	≥ 1.4	≥ 2.0	≥ 3.3
+ Likelihood Ratio	≤ 5.0	≤ 10.0	> 10.0	
– Likelihood Ratio	≤ 0.5	≤ 0.2	< 0.1	
Hazards Ratio	≥ 1.3	≥ 2.0	≥ 4.0	≥ 10.0

- Success in a grad program is difficult to define
  - Most commonly accepted indicator of academic success is GPA



	First-attempt Pass on the BOC exam	
	Yes	**No**
First-year gGPA ≥ 3.45	71	9
First-year gGPA < 3.45	19	20
Fisher's Exact Test (one-sided) $p < 0.001$		
Sn = 0.79 (95% CI: 0.69, 0.86)		Sp = 0.69 (95% CI: 0.51, 0.83)
OR = **8.30** (95% CI: 3.26, 21.16)		RFS = **1.82** (95% CI: 1.49, 2.23)
+LR = 2.54 (95% CI: 1.46, 4.42)		–LR = 0.36 (95% CI: 0.192, 0.489)

Potential Predictors for PMATP Success
(Major Categories Only)

- Academic Profile of Undergraduate Institution (APUI)
- Basic Carnegie classification categories
- Undergraduate institution size and setting
- Advanced math & science courses
- Number of adv. science courses
- Number of AT courses
- Adv. math, science, & AT courses
- uGPA
- GRE Scores
- Public-Private Institution
- Residency
  - In-state vs. Out-of-state

39 original variables investigated

Multicollinearity

- When 2 or more predictors in a regression model are highly linearly related
- Outcome parameter for multicollinearity is Tolerance & Variance Inflation Factor (VIF)
  - Tolerance values close to zero = multicollinearity
  - VIF = values of > 10 = multicollinearity
- Multicollinearity helped reduce the number of predictors from 39 to 9

Predictor Variables for PMATP Success

Origin Set of Predictors

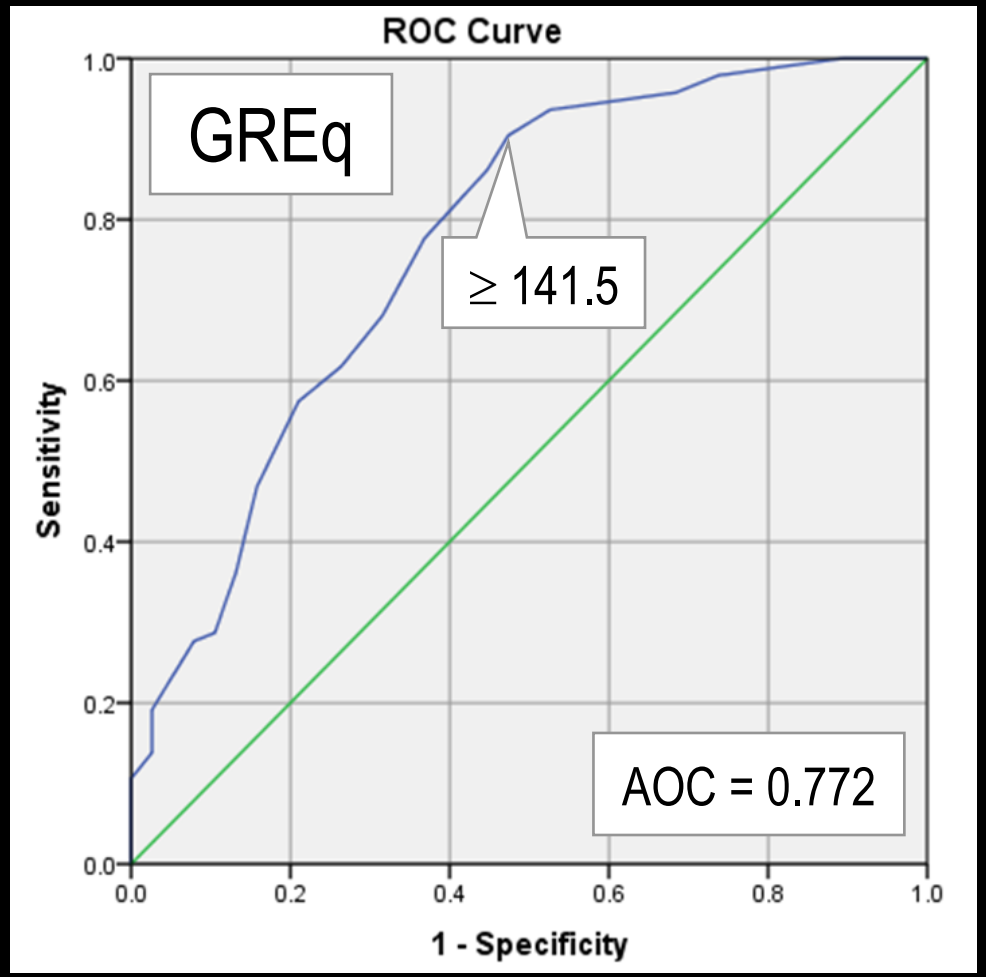
- Number of math & science courses
- Research Intensive = 1; Others = 0
- High APUI
- uGPA
- GREv
- GREq
- GREwr
- Physics: 1 = Yes; 0 = No
- Calculus: 1 = Yes; 0 = No

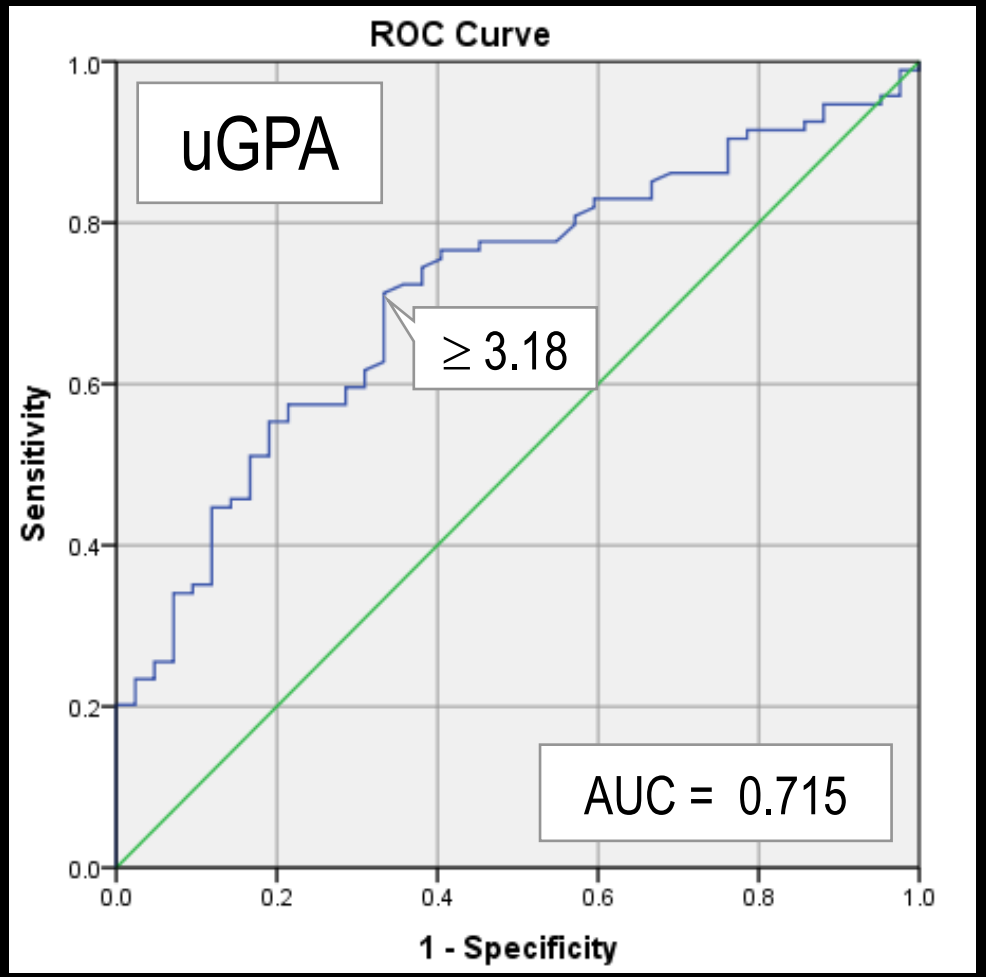
Final Set of Predictors

- uGPA
- GREq
- Calculus

- "Original Set of Predictors" is after multicollinearity analysis which were then entered into the logistic regression.
- The "Final Set of Predictors" were what predictors were left after logistic regression

ROC Curves for GREq & uGPA (for cut-pts)





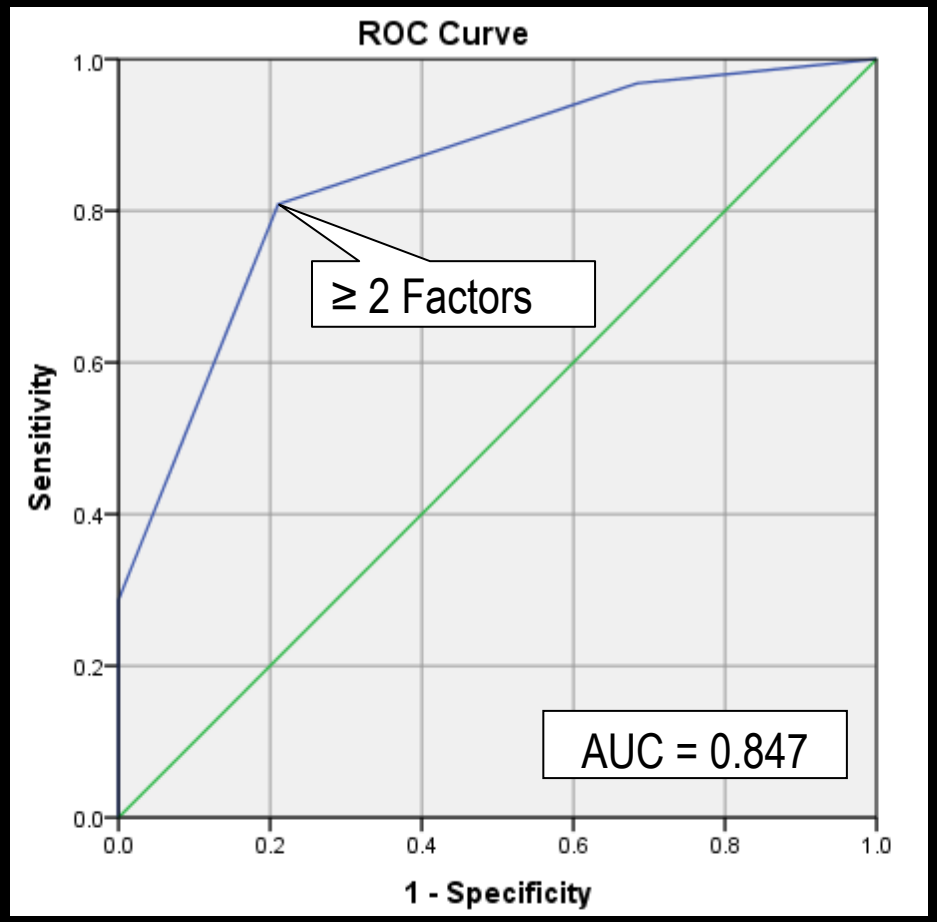
Tables for Individual Variables

Association	Small	Moderate	Large	Very Large
Odds Ratio	≥ 1.5	≥ 3.4	≥ 9.0	≥ 32.0
Relative Risk/Relative Frequency of Success	≥ 1.1	≥ 1.4	≥ 2.0	≥ 3.3
+ Likelihood Ratio	≤ 5.0	≤ 10.0	> 10.0	
– Likelihood Ratio	≤ 0.5	≤ 0.2	< 0.1	
Hazards Ratio	≥ 1.3	≥ 2.0	≥ 4.0	≥ 10.0

	1st Year gGPA ≥ 3.45	1st Year gGPA < 3.45		1st Year gGPA ≥ 3.45	1st Year gGPA < 3.45		1st Year gGPA ≥ 3.45	1st Year gGPA < 3.45
GREq = ≥ 141.5	85	18	**uGPA = ≥ 3.18**	68	15	**Took Calculus**	41	3
GREq = < 141.5	9	20	**uGPA = < 3.18**	26	27	**Did not take Calculus**	53	39
Fisher's Exact Test (one-sided) p < 0.001 for all 3 factors								
Sn (95% CI)	**0.90** (0.84, 0.95)			0.72 (0.63, 0.80)			0.44 (0.34, 0.54)	
Sp (95% CI)	0.53 (0.37, 0.68)			0.64 (0.49, 0.77)			**0.93** (0.81, 0.98)	
OR (95% CI)	**10.49** (4.11, 26.78)			**4.71** (2.17, 10.23)			**10.06** (2.90, 34.86)	
RFS (95% CI)	**2.66** (2.17, 3.26)			**1.67** (1.36, 2.05)			**1.62** (1.32, 1.98)	
+LR (95% CI)	1.91 (1.36, 2.86)			2.03 (1.33, 3.10)			**6.11** (2.00, 18.61)	
–LR (95% CI)	**0.182** (0.09, 0.36)			0.430 (0.29, 0.64)			0.607 (0.50, 0.74)	

Optimum Number of Predictors for PMATP Success

Association	Small	Moderate	Large	Very Large
Odds Ratio	≥ 1.5	≥ 3.4	≥ 9.0	≥ 32.0
Relative Risk/Relative Frequency of Success	≥ 1.1	≥ 1.4	≥ 2.0	≥ 3.3
+ Likelihood Ratio	≤ 5.0	≤ 10.0	> 10.0	
– Likelihood Ratio	≤ 0.5	≤ 0.2	< 0.1	
Hazards Ratio	≥ 1.3	≥ 2.0	≥ 4.0	≥ 10.0



uGPA ≥ 3.18; GREq ≥ 141.5; Student took calculus

	First-year gGPA ≥ 3.45	First-year gGPA ≥ 3.45
≥ 2 Factors	76	8
< 2 Factors	18	34

Fisher's Exact Test (one-sided): $p < 0.001$

Sn = 0.81 (95% CI: 0.72, 0.88)	Sp = 0.81 (95% CI: 0.67, 0.90)
OR = **17.94** (95% CI: 7.11, 45.29)	RFS = **2.61** (95% CI: 2.13, 3.20)
+LR = 4.25 (95% CI: 2.26, 7.98)	–LR = 0.237 (95% CI: 0.152, 0.367)

Specific Number of Factors for Prediction of PMATP Success

Number of Positive Factors	gGPA ≥ 3.45	gGPA < 3.45	Total	%	% above/ below cut point
0	3	16	19	16%	18/52 = 35%
1	15	18	33	45%	
2	49	9	57	86%	76/84 = 91%
3	27	0	27	100%	
Total	94	42	136	70%	

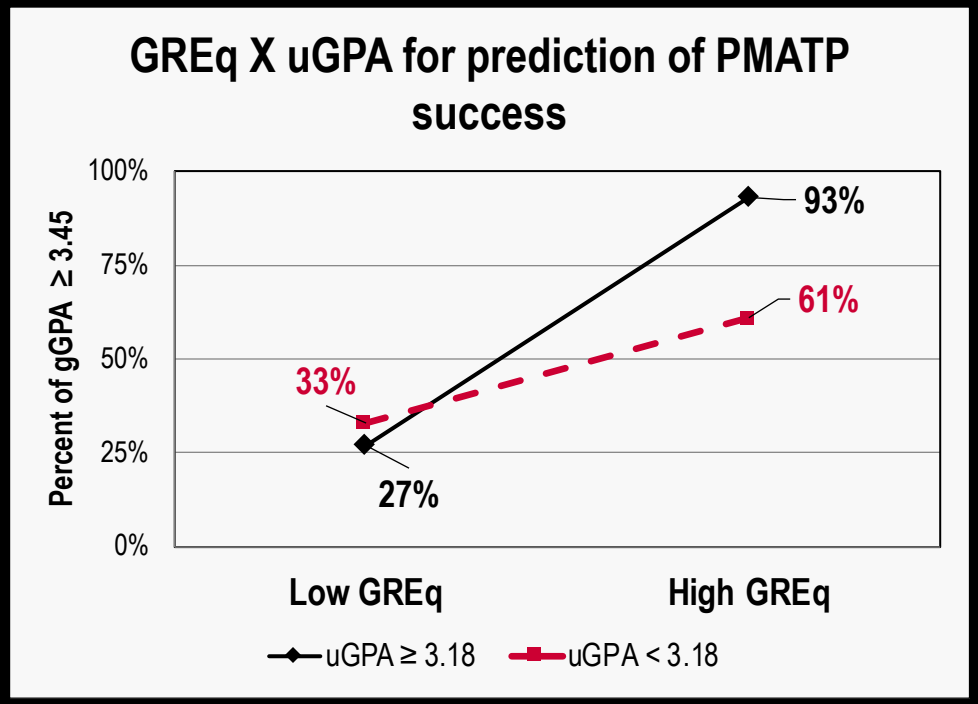
0.91 / 0.35 = 2.6 (RFS)

Interaction Effects

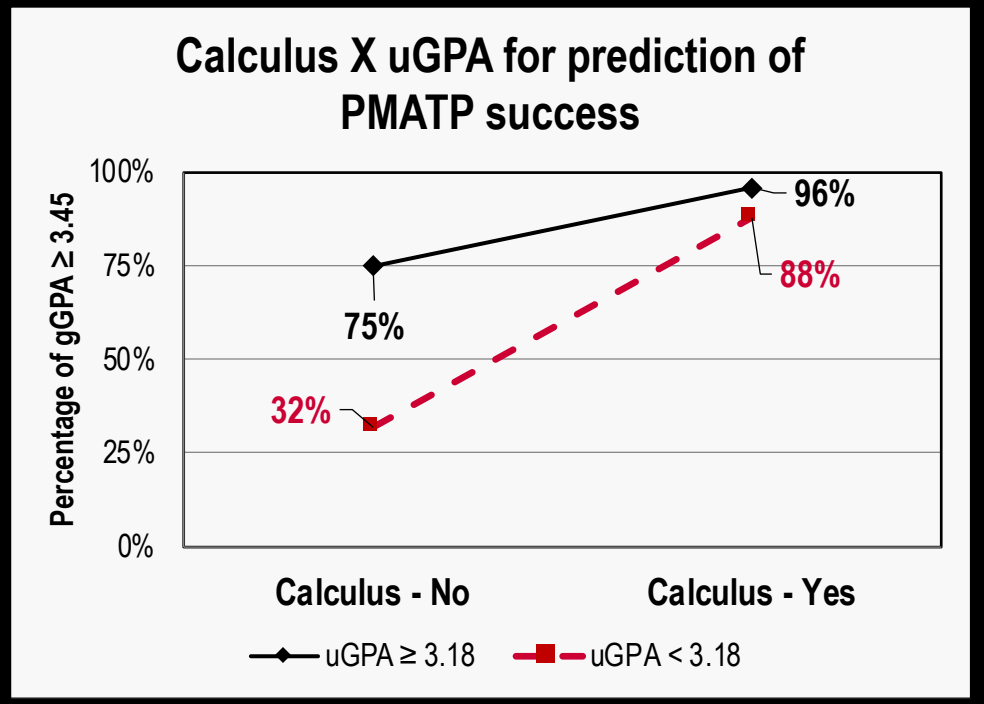
Comparison of Odds Ratios for Predictor Variables

	Univariable OR	Multivariable Adj OR
uGPA	4.71 (95% CI: 2.17, 10.23)	7.62 (95% CI: 2.63, 22.13)
GREq	10.49 (95% CI: 4.11, 26.78)	7.68 (95% CI: 2.48, 23.76)
Calculus	10.06 (95% CI: 2.90, 34.86)	11.77 (95% CI: 2.66, 52.11)

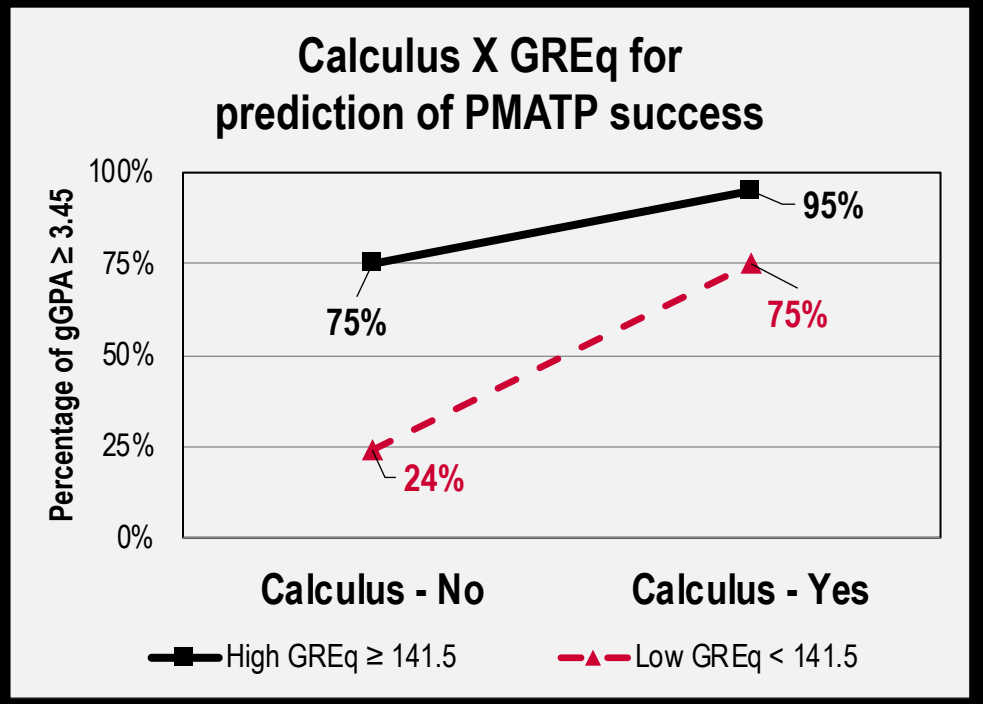
- An interaction btw uGPA &GREq is suggested by the differences btw the univariable OR & the corresponding multivariable adjusted OR
- Relatively little change btw the univariable OR & the corresponding multivariable adj. OR for taking calculus


GREq X uGPA for prediction of PMATP success
Percent of gGPA ≥ 3.45
100%
75%
50%
25%
0%
93%
61%
33%
27%
Low GREq
High GREq
uGPA ≥ 3.18
uGPA < 3.18

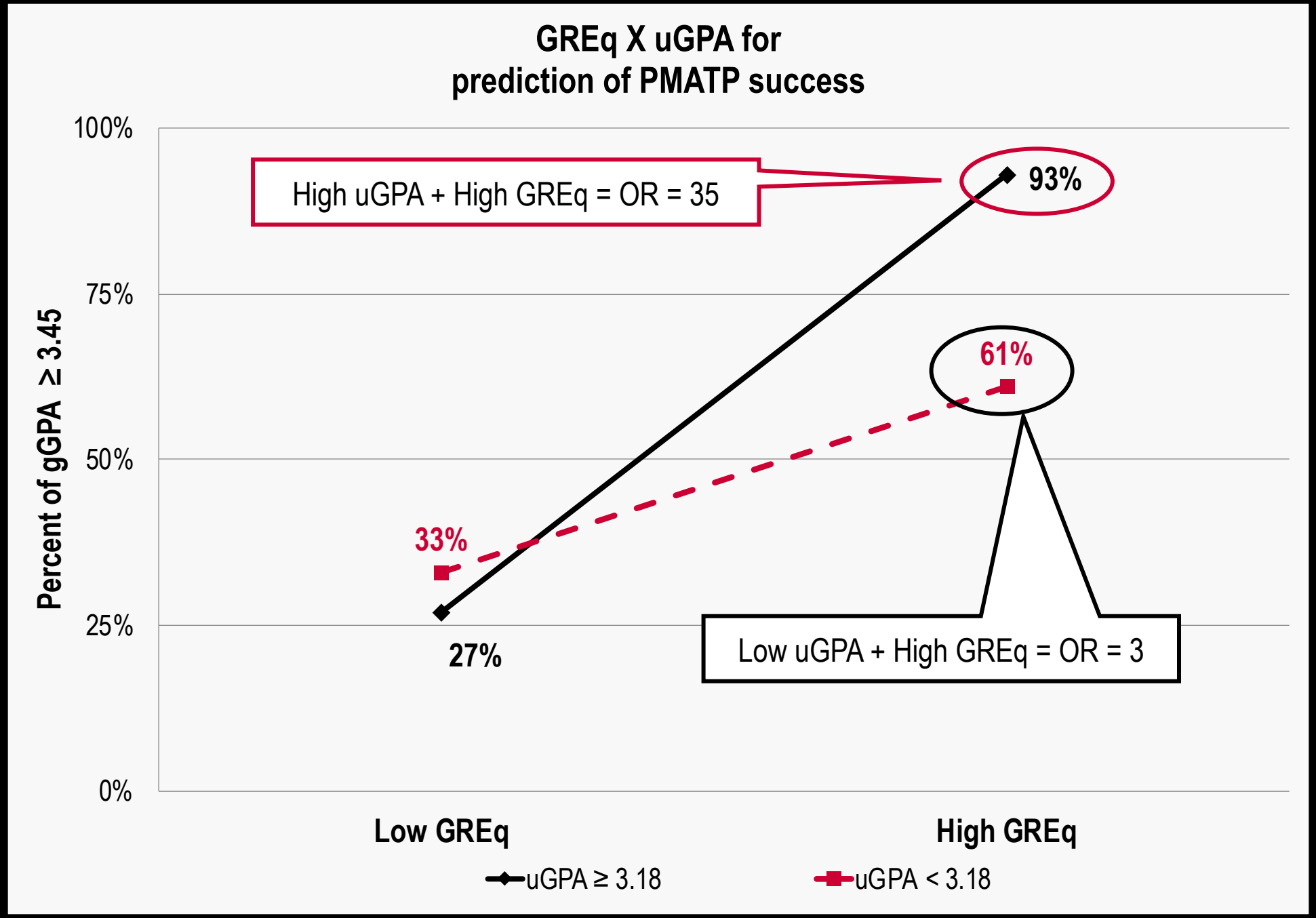
A.


Calculus X uGPA for prediction of PMATP success
Percentage of gGPA ≥ 3.45
100%
75%
50%
25%
0%
96%
88%
75%
32%
Calculus - No
Calculus - Yes
uGPA ≥ 3.18
uGPA < 3.18

B.

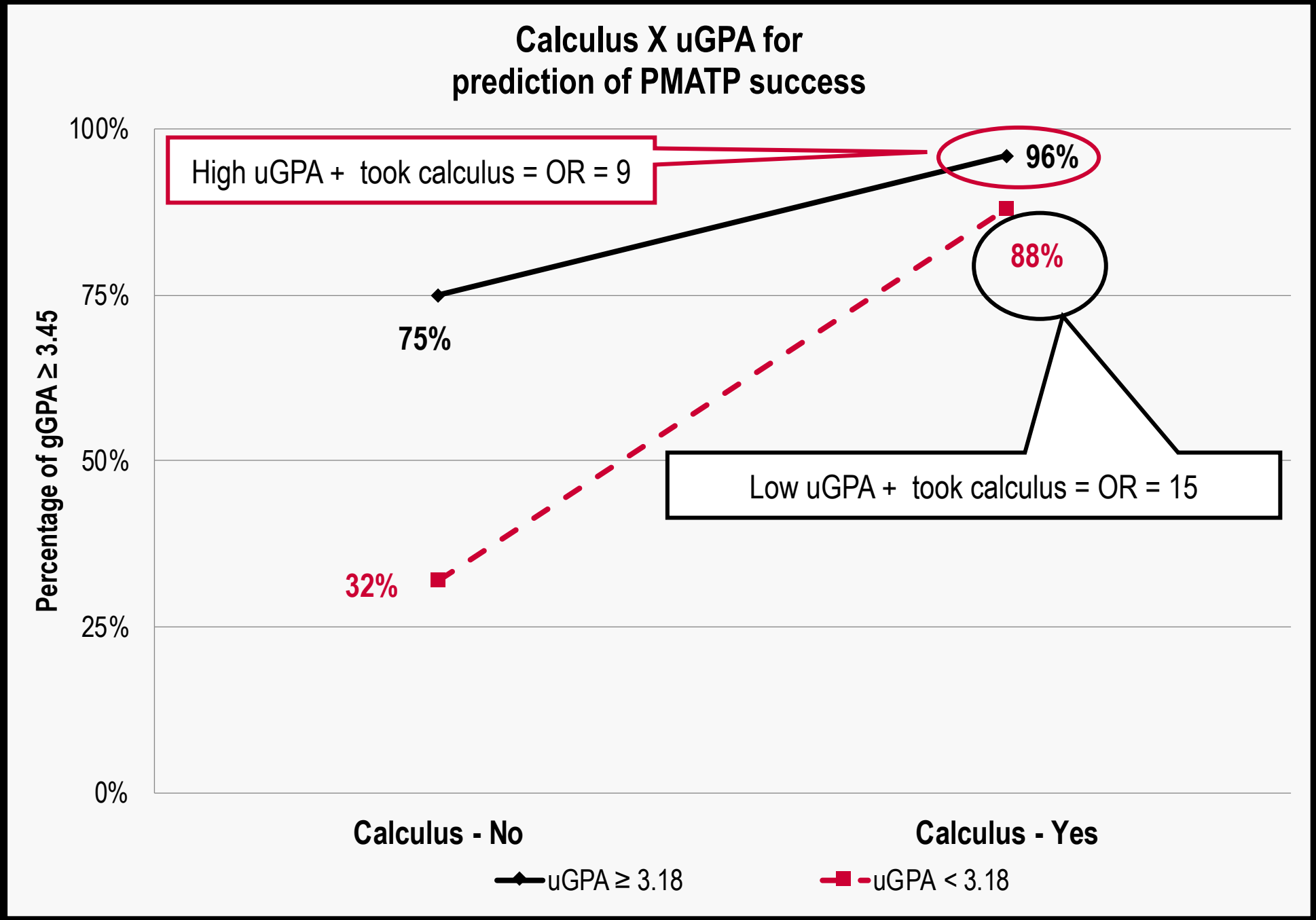

Calculus X GREq for prediction of PMATP success
Percentage of gGPA ≥ 3.45
100%
75%
50%
25%
0%
95%
75%
75%
24%
Calculus - No
Calculus - Yes
High GREq ≥ 141.5
Low GREq < 141.5

C.


GREq X uGPA for
prediction of PMATP success
100%
75%
50%
25%
0%
Percent of gGPA ≥ 3.45
High uGPA + High GREq = OR = 35
93%
61%
33%
27%
Low uGPA + High GREq = OR = 3
Low GREq
High GREq
uGPA ≥ 3.18
uGPA < 3.18

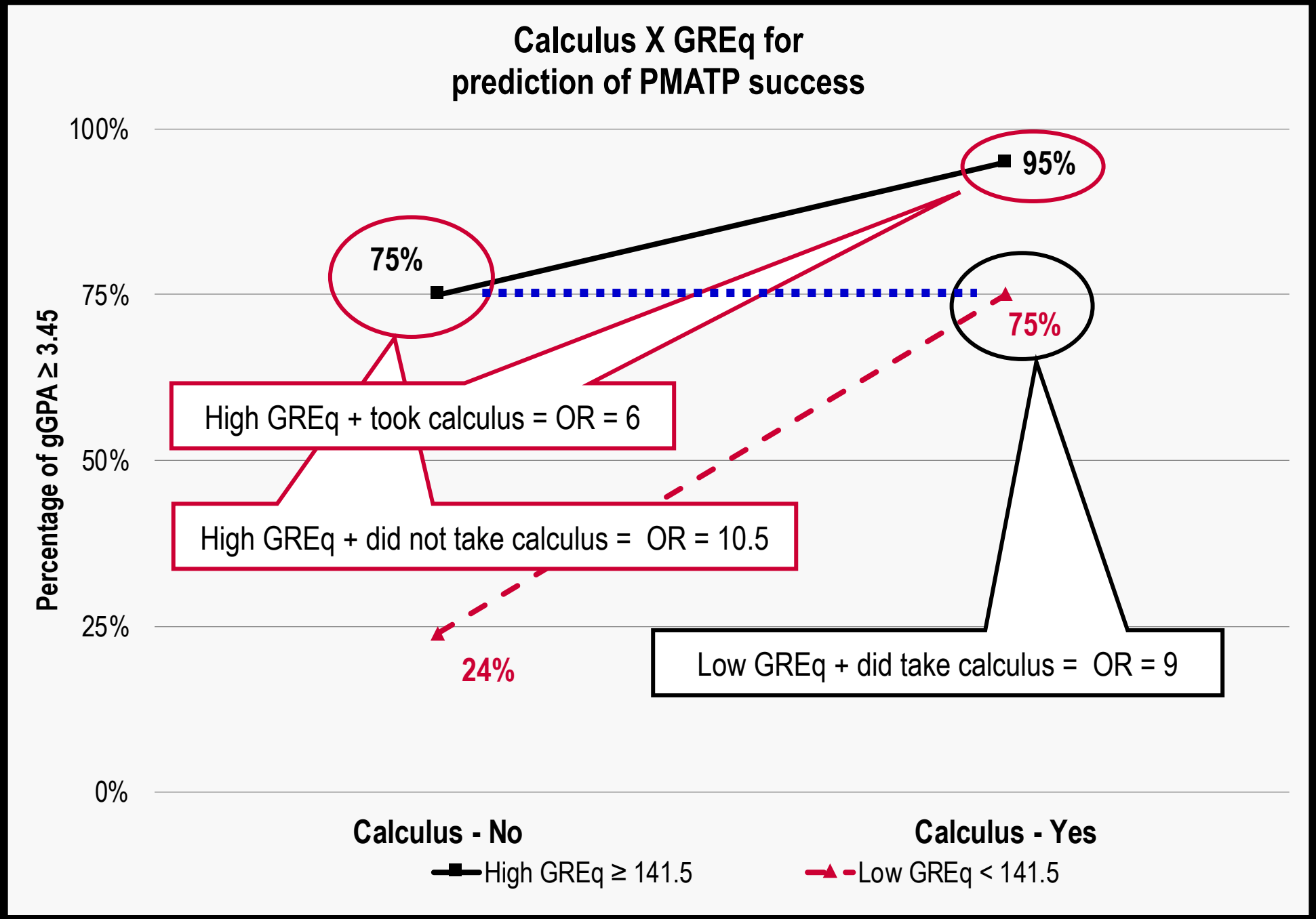
Effect of GREq X uGPA & PMATP Success

- Controlling for uGPA strata (≥ 3.18 vs. < 3.18):
  - Relationship btw GREq and being successful in the PMATP was examined
    - Mantel-Haenszel OR_{est} = 6.5 (95% CI: 2.59, 16.52)
  - There is statistically significant association between GREq and PMATP success
    - Mantel-Haenszel $\chi^2(1) = 18.62$; ($p < 0.001$)
  - The null hypothesis for the Breslow-Day test assumes that the ORs for GREq predicting PMATP success is equivalent for uGPA strata
  - Breslow-Day test for homogeneity found the ORs to be significantly different for the two strata of uGPA
    - Breslow-Day $\chi^2(1) = 6.05$; ($p = 0.014$)


Calculus X uGPA for
prediction of PMATP success
100%
75%
50%
25%
0%
Percentage of gGPA ≥ 3.45
High uGPA + took calculus = OR = 9
96%
75%
88%
32%
Low uGPA + took calculus = OR = 15
Calculus - No
Calculus - Yes
uGPA ≥ 3.18
uGPA < 3.18

Effect of Calculus X uGPA & PMATP Success

- Controlling for uGPA strata (≥ 3.18 vs. < 3.18):
  - Relationship btw taking calculus and being successful in the PMATP was examined
    - Mantel-Haenszel OR_{est} = 11.8 (95% CI: 3.71, 44.12)
  - There is statistically significant association between taking calculus and PMATP success
    - Mantel-Haenszel $\chi^2(1) = 16.76$; ($p < 0.001$)
  - The null hypothesis for the Breslow-Day test assumes that the ORs for taking calculus predicting PMATP success is equivalent for uGPA strata
  - The Breslow-Day test for homogeneity found the ORs to not be significantly different from one another
    - Breslow-Day $\chi^2(1) = 0.12$; ($p = 0.730$)

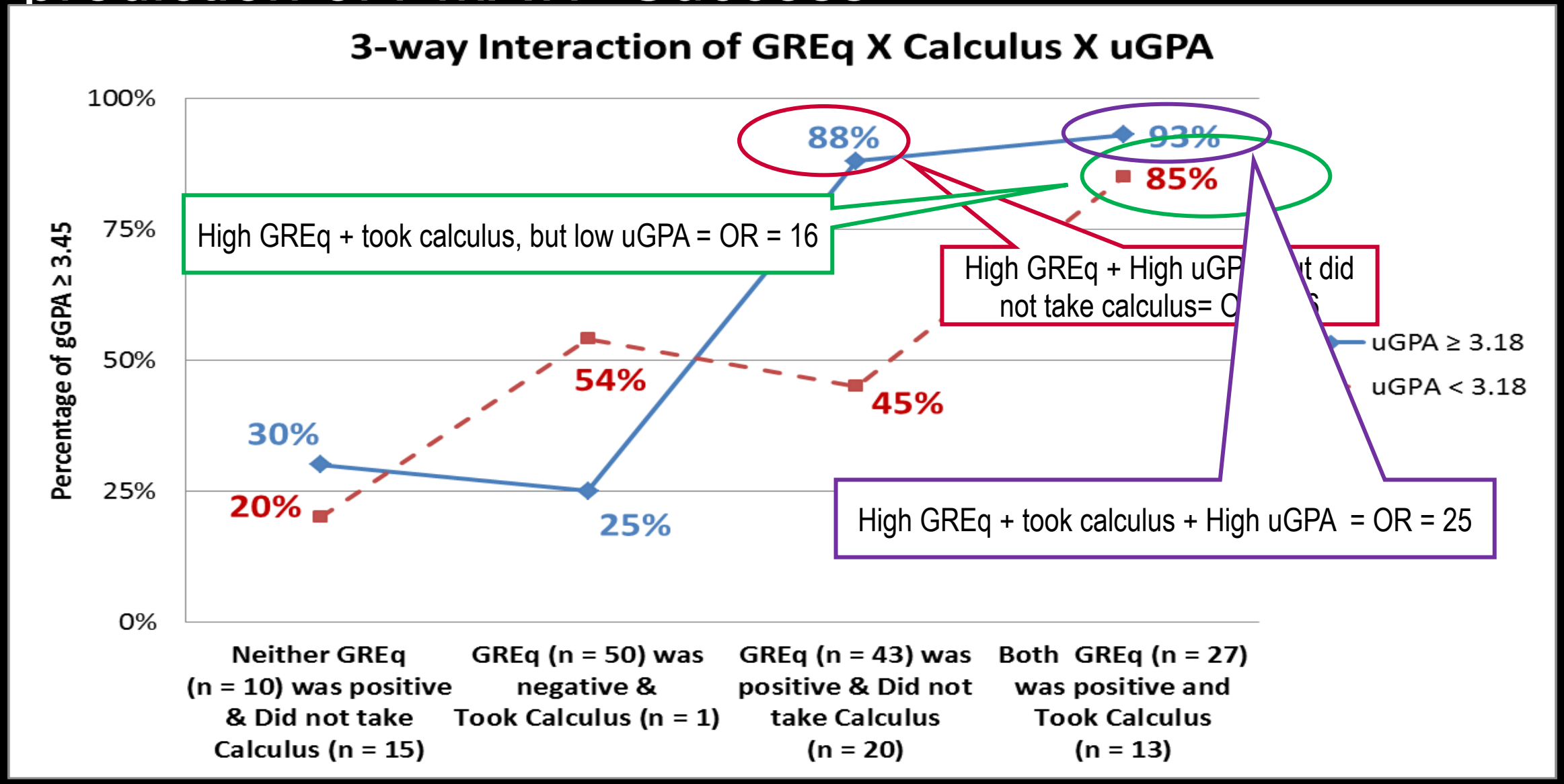

Calculus X GREq for
prediction of PMATP success
100%
75%
50%
25%
0%
Percentage of gGPA ≥ 3.45
95%
75%
75%
High GREq + took calculus = OR = 6
High GREq + did not take calculus = OR = 10.5
24%
Low GREq + did take calculus = OR = 9
Calculus - No
Calculus - Yes
High GREq ≥ 141.5
Low GREq < 141.5

Effect of Calculus X GREq & PMATP Success

- Controlling for GREq (≥ 141.5 vs. < 141.5):
  - Relationship btw taking calculus and being successful in the PMATP was examined
    - Mantel-Haenszel OR_{est} = 10 (95% CI: 3.29, 24.49)
  - There is statistically significant association between taking calculus and PMATP success
    - Mantel-Haenszel $\chi^2(1) = 18.85$; $p < 0.001$)
  - The null hypothesis for the Breslow-Day test assumes that the ORs for taking calculus predicting PMATP success is equivalent for GREq strata
  - The Breslow-Day test for homogeneity found the ORs to not be significantly different from one another
    - Breslow-Day $\chi^2(1) = 0.07$; ($p = 0.791$)

Three-way Interactions

3-way interaction of GREq X Calculus X uGPA for prediction of PMATP Success



Are you here yet?

THANK YOU!!!

Scott L. Bruce, EdD, ATC
Associate Professor/Clinical Education Coordinator
Arkansas State University

sbruce@astate.edu
423-240-1709

Jeremiah 29:11