Models to Predict the Risk of Infection After Surgery in Pancreatic Cancer Patients

Jolene Ranek
Brooke Bell
Jeffrey Antunes
Mentor: Dr. Brian Smith

"In 2014, it's estimated over 46,000 people in the U.S. will be diagnosed with pancreatic cancer…"

- 2 glands: exocrine and endocrine
- Enzymes, islets
- Unknown cause for cancer
- Biopsy
- Stages 1-4
  - 1 – cancer is limited to pancreas
  - 2- spread to lymph nodes
  - 3- spread to blood vessels
  - 4- metastasized to nearby organs

- *"Cancer of the pancreas barely makes the top 10 most common cancers in the U.S. However, pancreatic cancer's tendency to spread silently before diagnosis makes it the fourth deadliest cancer diagnosis with more than 39,000 people expected to die of the disease in 2014."*

- Asymptomatic until the later stages
- 1-year survival rate is 20%, and the five-year rate is 6%

http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.891917/k.5123/Prognosis_of_Pancreatic_Cancer.htm

Overview of Treatment

- Surgery in combination with chemotherapy and radiation
- Surgery may be done to remove the cancerous tumor or lessen symptoms
  - Whipple - most common operation to remove a cancer of the exocrine pancreas; surgeon removes the head of the pancreas and nearby structures
  - Distal - surgeon removes only the tail of the pancreas and the spleen
- Chemotherapy and radiation are often given together- used to slow growth of pancreatic cancer
- In those cases where resection can be performed, the average survival rate is 18 to 20 months

http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.891917/k.5123/Prognosis_of_Pancreatic_Cancer.htm

Study Objective

- Create predictive models that can be used to assess the risk of surgical complications

Overview of Variables

- American College of Surgeons National Surgical Quality Improvement Program (NSQIP)
  - 150+ variables for 23,000+ patients
  - Preoperative risk factors
  - Intraoperative variables

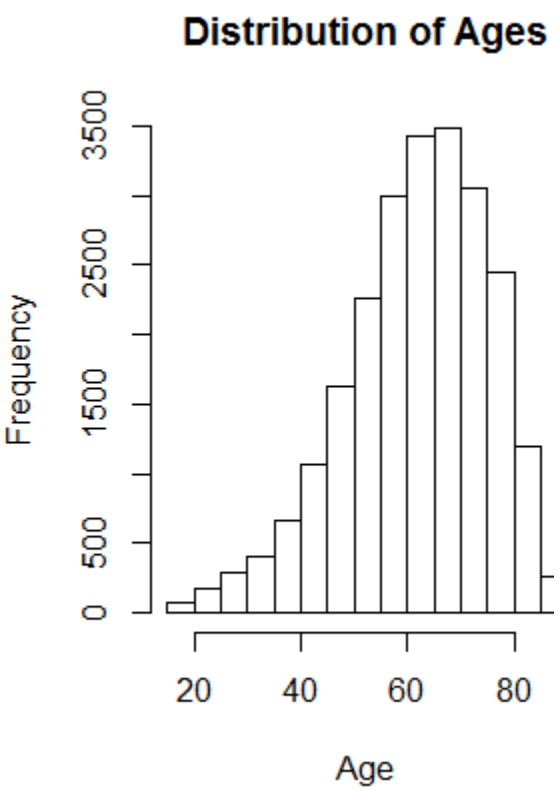
Overview of Variables

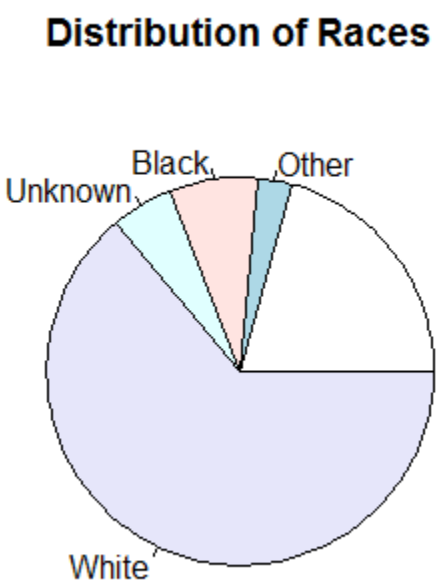
- Examined 50+ variables
  - Continuous and categorical
  - Ex: Age, height, race, prior operation in past 30 days, history of transient ischemic attacks, type of procedure
- Examined 3 separate outcomes
  - Surgical Site Infections
  - Data accounts for infections that occurred within 30 days after surgery

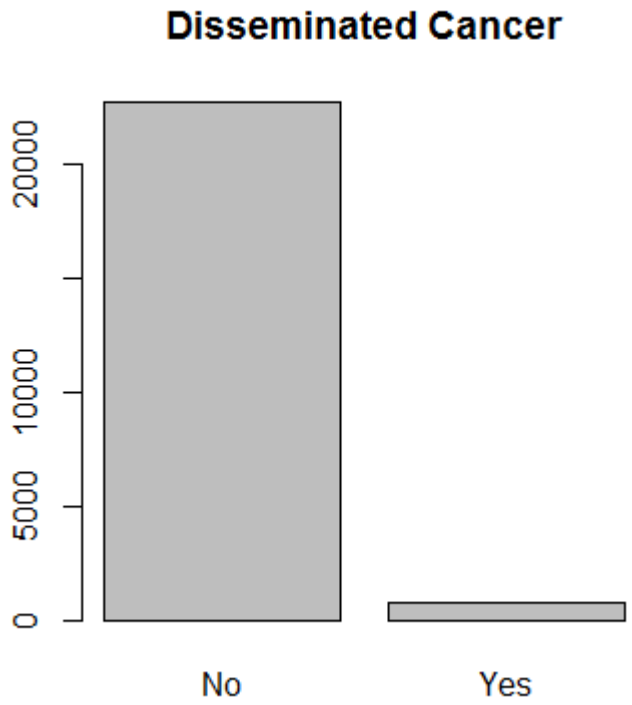
Surgery Outcomes

- Surgical Site Infections (SSI)
  - Superficial Incisional SSI: skin or subcutaneous tissue
  - Deep Incisional SSI: deep soft tissue
  - Organ Space SSI: organs or spaces

Summary Statistics







Summary Statistics

OUTCOME	NO INFECTION	INFECTION	PERCENT (%)
Superficial Incisional SSI	21563	1934	8.23
Deep Incisional SSI	23008	489	2.08
Organ Space SSI	21076	2421	10.30

Analysis Methods

- Univariate Analysis
  - Conducted univariate analysis for each variable to identify the most statistically significant based off of their p-values
  - Used these variables as an initial starting point for our multivariate model

Analysis Methods

- Multivariate Analysis
  - Built a multivariate regression model to predict outcomes
  - Identified a set of variables that carry the most information for predicting the outcome
  - Used to make a logistic regression model

Logistic Regression

$$Y \sim \text{Binomial}(1, \pi(x))$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_\circ + \beta_1 x_1 + \ldots + \beta_n x_n$$

- Y is coded as 1 for infection and 0 for no infection
- Assumed to have a binomial distribution
- π is the probability of the infection modeled as a predictor variable x
- β is the effect of predictor on infection probability

Training and Test Set of Data

- To construct our models, we took the data and randomly divided the patients into two equal groups
  - Training group – the group we used to construct our final models
  - Test group – an entirely independent group that we applied our model to assess its predictive performance
  - We used the area under the ROC curve to evaluate the predictive performance of the models

Surgery Outcome: Superficial Incisional SSI

http://blog.sisfirst.com/blog/bid/147767/What-s-the-Secret-to-a-High-Performing-Surgical-Department

Step-wise Model Selection

- Applied step-wise model selection on the variables that were statistically significant in the univariate model at the 0.5 alpha level
- Created a model by backward step-wise selection
- Final model had 10 variables

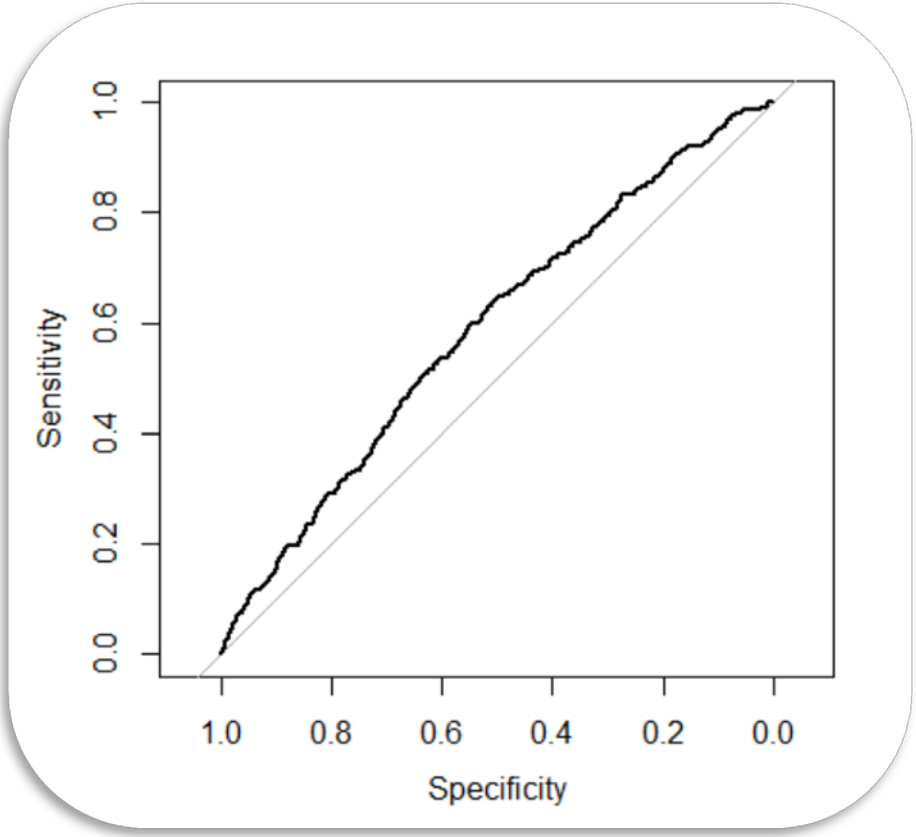
Significant Variables

- Patient characteristics that influence the risk of infection in superficial tissue include:
  - Age
  - Weight
  - Previous angioplasty (PCI)
  - Pre-op bun
  - Pre-op serum albumin
  - Pre-op phosphatase alkaline
  - Pre-op platelet count
  - Pre-op normalized PT values
  - Radio therapy
  - Bleeding disorder

Predictive Power

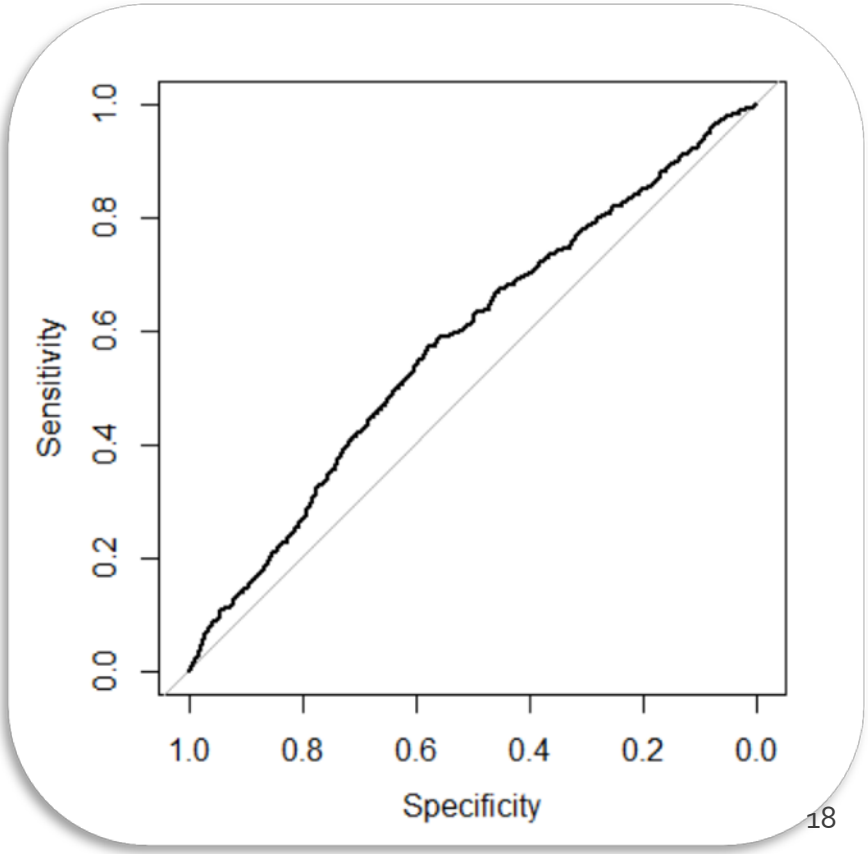
ROC curve for training set

Area under the curve is 0.5888



ROC curve for test set

Area under the curve is 0.5783



Odds Ratios

Variable	SD * Normal Levels **	Odds Ratios	Lower Bounded 95 % CI	Upper Bounded 95 % CI
Age	13.40 *	1.11	2.73	3.39
Weight	43.36 *	1.17	2.92	3.55
Previous angioplasty (PCI)	1	1.32	2.60	5.41
Pre-op bun	13 **	0.84	1.91	2.79
Pre-op serum albumin	2 **	0.76	1.56	2.94
Pre-op phosphatase alkaline	103 **	1.07	2.74	3.08
Pre-op platelet count	250 **	1.30	2.91	4.66
Pre-op normalized PT values	0.30 **	0.78	1.73	2.75
Radio therapy - No	1	1	-	-
Radio therapy - Yes	1	1.57	2.99	7.74
Bleeding Disorder - No	1	1	-	-
Bleeding disorder - Yes	1	1.52	2.89	7.29

Surgery Outcome: Organ Space SSI

http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.887625/k.9A08/Pancreatic_Cancer.htm

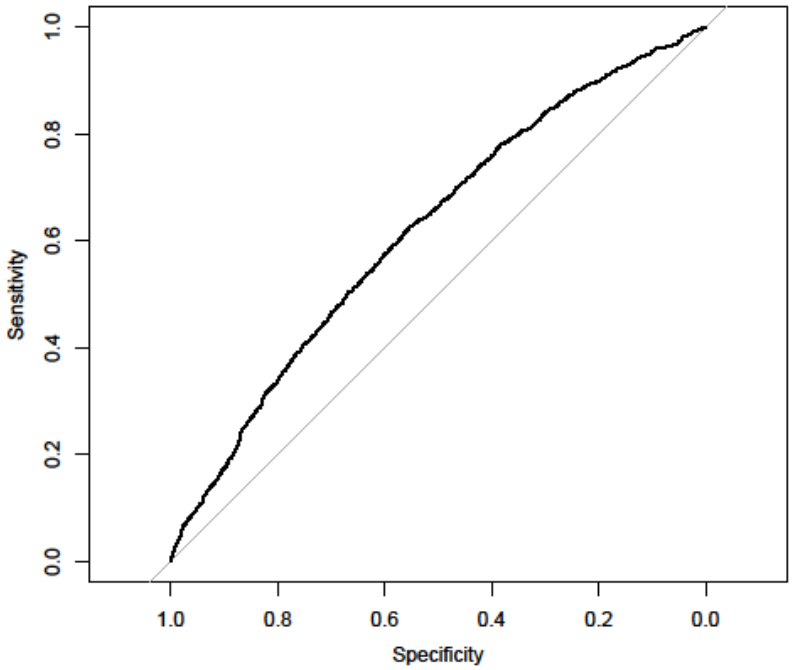
Significant Variables

- Initial model had 32 variables and after implementing stepwise model selection, final model had 15 variables
  - Weight
  - Sex
  - CPT
  - Dyspnea
  - Pre – operative serum albumin
  - Pre – operative alkaline phosphatase
  - Pre – operative hematocrit
  - Disseminated cancer
  - Weight loss
  - Bleeding disorder
  - Emergency case
  - Height
  - Diabetes
  - Previous cardiac surgery
  - Pre – operative serum sodium

Predictive Power

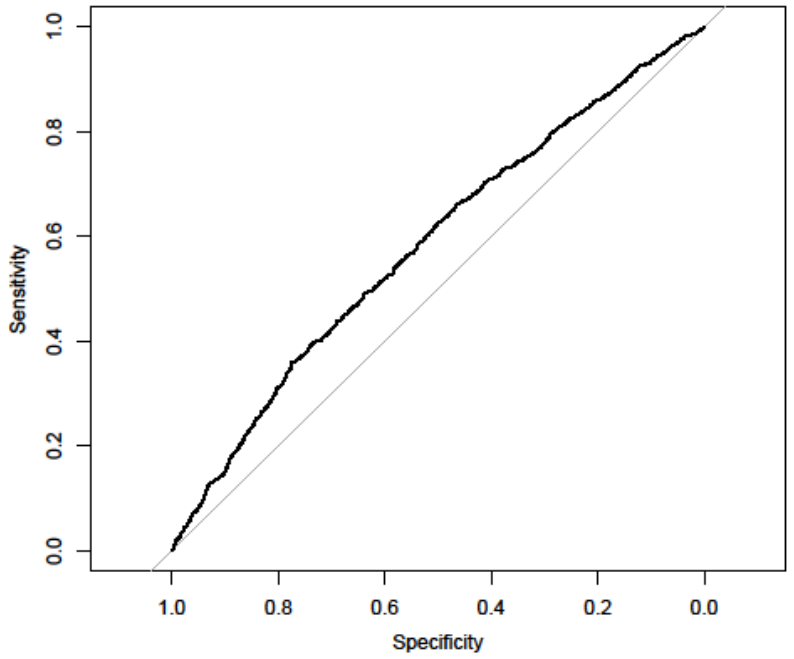
ROC curve for training set

Area under the curve (AUC) is 0.6195



ROC curve for test set

Area under the curve (AUC) is 0.5836



Odds Ratios

Variable	SD* Normal levels**	Odds Ratio	Lower Bounded 95% CI	Upper Bounded 95% CI
Dyspnea Moderate Exertion	1	2.73	5.36	43.59
Dyspnea - No	1	1.70	1.96	15.43
Dyspnea - At Rest	1	1	-	-
Pre-operative serum albumin	2**	0.62	1.46	2.38
Disseminated cancer - Yes	1	1.67	3.90	7.17
Disseminated cancer - No	1	1	-	-
Previous cardiac surgery - Yes	1	0.66	1.32	2.85
Previous cardiac surgery - No	1	1	-	-

Surgery Outcome: Deep Incisional SSI

http://www.maquet.com/int/product/discipline/cardiac-surgery

Significant Variables

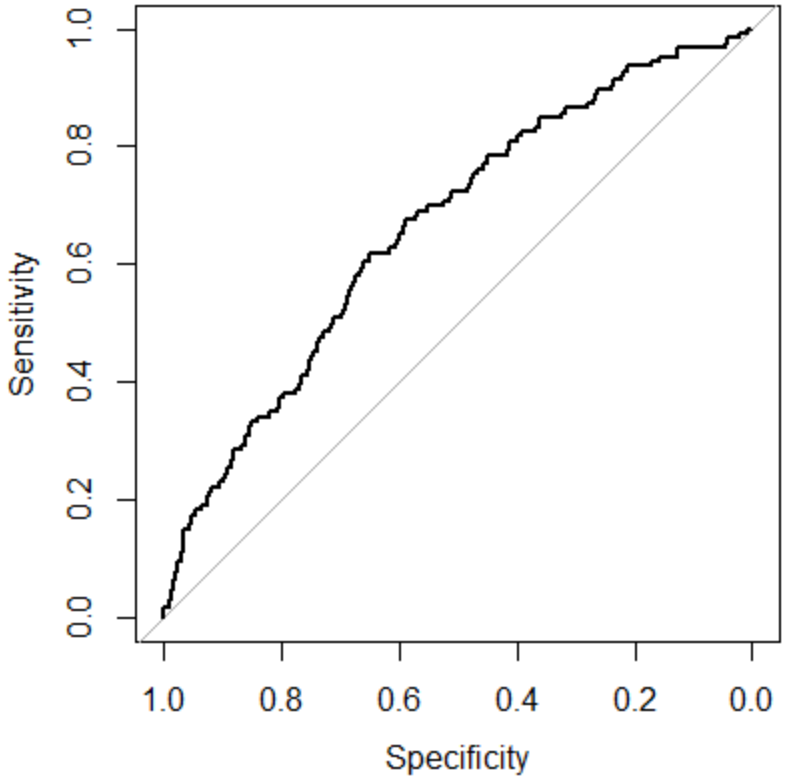
Variable Name	P-value
Height	0.0174
Weight	< 0.001
Sex	0.0107
Ethnicity Hispanic	0.0322
CPT code	< 0.001
Pre-operative serum albumin	0.0029
Pre-operative bilirubin	0.0213
Pre-operative PT	0.0396
Probability of morbidity	< 0.001
ASA classification	0.0146
Bleeding disorder	0.0217

Model Selection

- Backward selection
  - Alpha level = 0.05
  - Began with all variables of interest
  - Dropped least significant variable
  - Continued process until all remaining variables were statistically significant at $\alpha = 0.10$
- Final model included 6 variables
  - Weight, Hispanic ethnicity, type of surgery (CPT), pre-operative serum albumin, pre-operative PT, bleeding disorders

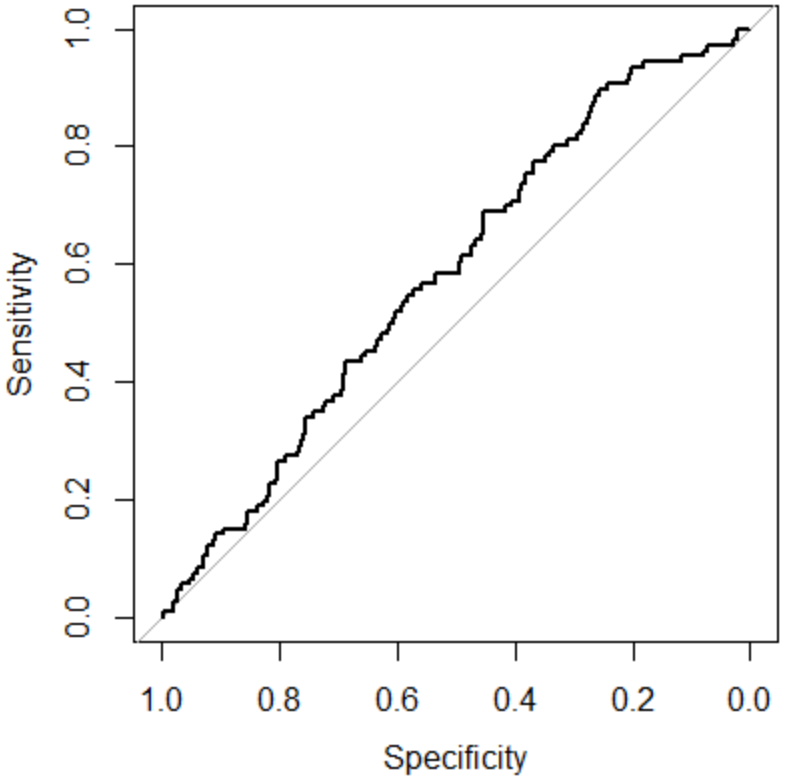
Predictive Power

Training set of data



Area under the curve
(AUC) = 0.6618

Test set of data



Area under the curve
(AUC) = 0.5814

Odds Ratios

Variable	Odds Ratio	Lower Bounded 95% CI	Upper Bounded 95% CI
Weight	1.32	1.12	1.18
Hispanic Ethnicity – Not Reported	1	-	-
Hispanic Ethnicity - No	1.61	1.05	2.49
Ethnicity Hispanic - Yes	1.25	0.37	4.19
CPT - Other	1	-	-
CPT - Distal	1.87	0.55	6.39
CPT - Whipple	3.57	0.49	25.84
CPT - Total	2.47	0.22	27.64
Pre-operative serum albumin	0.74	0.57	0.96
Pre-operative PT	0.89	0.81	0.98
Bleeding disorders - Yes	2.29	1.13	4.64
Bleeding disorders - No	1	-	-

Conclusion

- A good predictive model has an AUC of at least 0.75 or higher
  - Shows that our models poorly predict the risk of infection
  - Although these models do not predict risk well, the results can still be used for future research
  - Possibility to combine all forms of infection and create a model based off of that

Thank You

Acknowledgements

Dr. Brian Smith
Dr. Gideon Zamba
Terry Kirk
Joe Moen & Melissa Pugh
John VanBuren
Biostatistics Graduate Students
Biostatistics Faculty and Staff
University of Iowa College of Public Health
National Institute of Health

Questions