Jorgensen, S., Fichten, C.S., & Havel, A. (2008, October). Predicting student attrition - How helpful are surveys? Presentation at the Canadian Institutional Research and Planning Association (CIRPA) annual convention, Quebec. Available in the Conference Proceedings, Number 23. http://www.cirpa-acpri.ca/quebec2008/webdav/site/acpri-cirpa2008/shared/Fichiers/Presentations/MA-5.pdf

Predicting Student Attrition

How Helpful are Surveys?

Predicting Student Attrition How Helpful are Surveys?

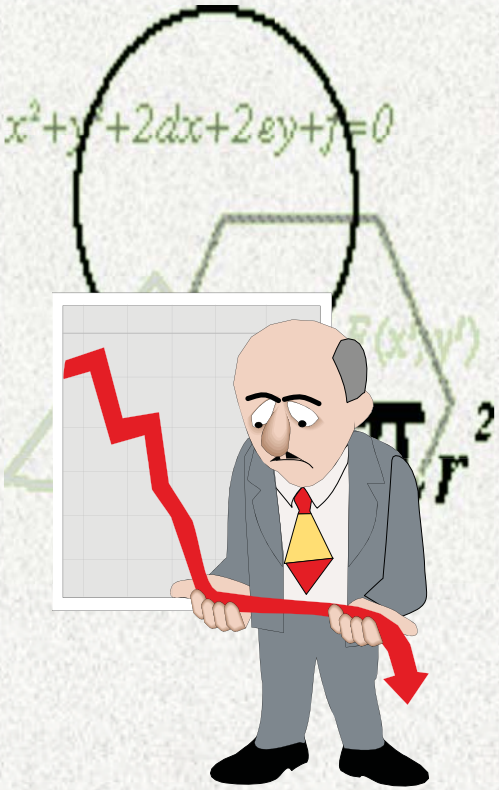
Shirley Jorgensen, Catherine Fichten, Alice Havel
Dawson College, Montreal Quebec
The Adaptech Research Network, Dawson College

Study made possible by:

- Funding received from
  - Canadian Council on Learning
  - PAREA (Quebec)
- And support from Dawson College

The Trouble With Surveys

- Partial coverage
- Low response rates
- Cost of administration
- Time to analyze data
- Non-response error



Surveys – Are the Costs Justified

Does survey data improve the ability to predict attrition enough to justify the costs?

Seven Models Tested

	Model
1	High school grade (HSG)
2	Records variables (8)
3	Records variables (8) & HSG
4	Survey variables (9)
5	Survey variables (9) & HSG
6	Records variables (8) & Survey variables (9)
7	Records variables (8) & Survey Variables (9) & HSG

Variables – From Records

- High school grade
- Country of birth
- Language
- English placement test (level)
- Sector of enrolment (2 or 3 year)
- Age
- Sex
- Disability
- Median income (Post code)

Variables From Surveys –Demographic etc

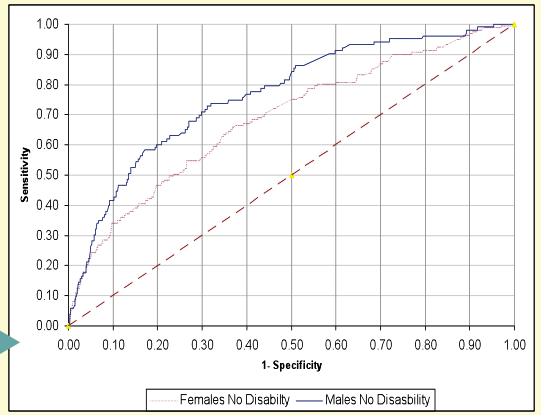
- Level of motivation
- First choice program
- Degree aspirations
- First generation college student
- COB – Mother
- COB – Father
- Anticipated hrs paid work
- Hours in study – last yr of study
- Anticipated hours of study - college

Tools Used

- Binary Logistic regression
- Nagelkerke R2
- Probability of dropout
- Coefficient(s) to calculate probability for new sample
- Classification matrix (for different cutoffs)

Observed	R	A	% C
Ret	553	273	66.9
Att	72	60	45.5
			64.0

- ROC curves (Area)
- Plots Sensitivity vs false positive rate for each cutoff (probability)



Classification Matrix

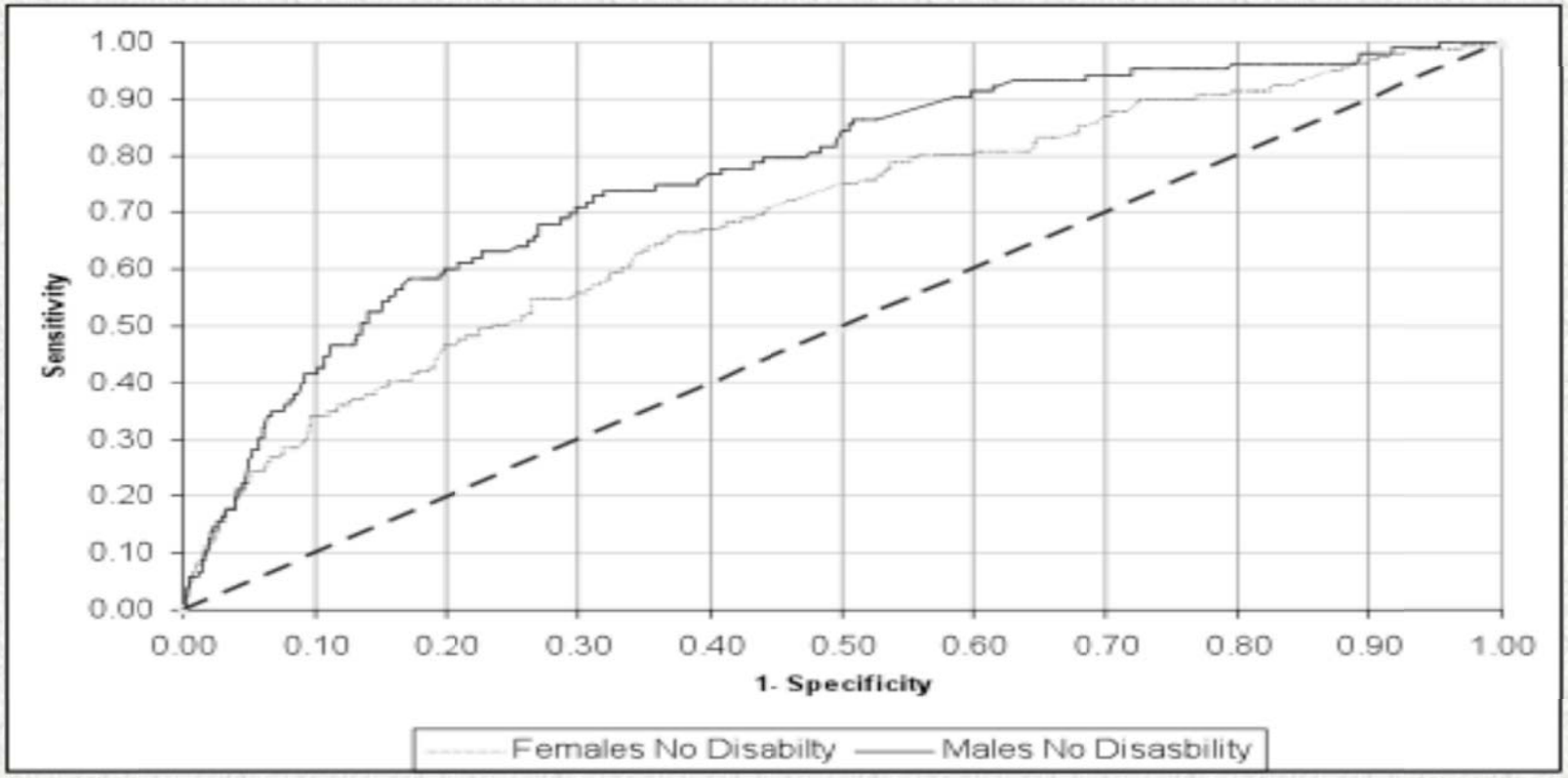
Cutoff = .4	Predicted				
Observed	Retention	Attrition	% Correct		
Retention	553	273	66.9	Specificity	1- Specificity (False Positive)
Attrition	72	60	45.5	Sensitivity	1 – Sensitivity (False Negative)
Overall Percentage			64.0		

ROC Data

ROC Graph Output			Calculate			
Positive if Greater Than or Equal To(a) (Cutoff or Probability)	Sensitivity	1 – Specificity (False Positive)	*Predicted Attrition (Number)*	*False Positive (Number)*	*Total Predicted Attrition*	*% Correct*
. .	. .	. .	. .	. .	. .	. .
0.155	0.623	0.364	*415*	*1274*	*1689*	*24.6%*
0.156	0.620	0.357	*413*	*1249*	*1662*	*24.8%*
0.157	0.615	0.351	*410*	*1229*	*1639*	*25.0%*
0.158	0.608	0.346	*405*	*1210*	*1615*	*25.1%*
0.159	0.604	0.340	*402*	*1190*	*1592*	*25.3%*
0.160	0.594	0.333	*396*	*1166*	*1562*	*25.3%*
0.161	0.587	0.328	*391*	*1147*	*1538*	*25.4%*
0.162	0.582	0.322	*388*	*1128*	*1516*	*25.6%*
0.164	0.577	0.314	*385*	*1099*	*1484*	*25.9%*
0.165	0.571	0.310	*380*	*1084*	*1464*	*26.0%*
0.166	0.561	0.303	*374*	*1060*	*1434*	*26.1%*
0.167	0.548	0.297	*365*	*1038*	*1403*	*26.0%*
0.168	0.540	0.292	*360*	*1020*	*1380*	*26.1%*

Receiver Operating Characteristic Curve (ROC)

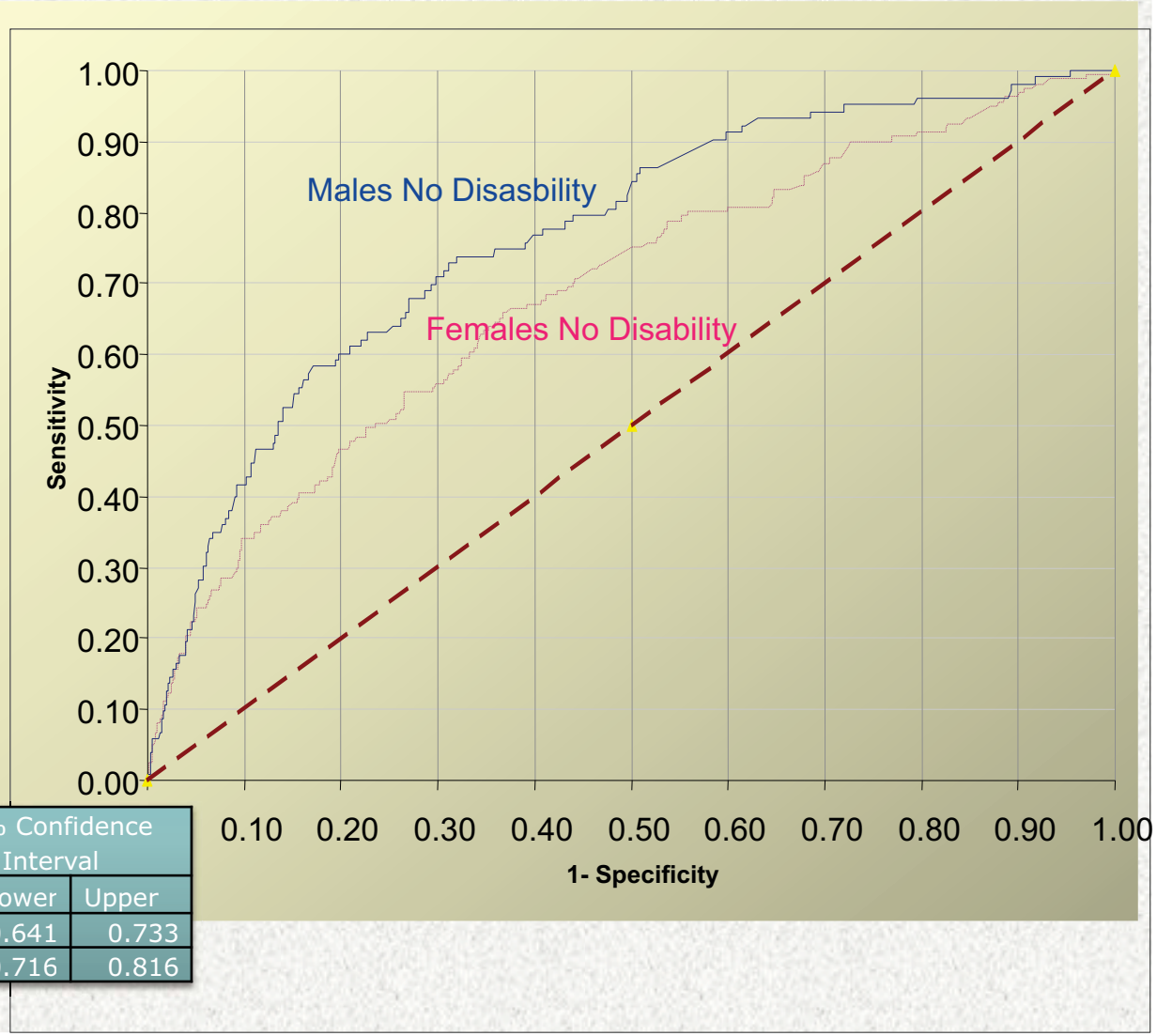
Two-dimensional depiction of classifier performance. ROC Accuracy Ratio, a common technique for judging the accuracy of default probability models.



Area Under the ROC Curve (AUC)

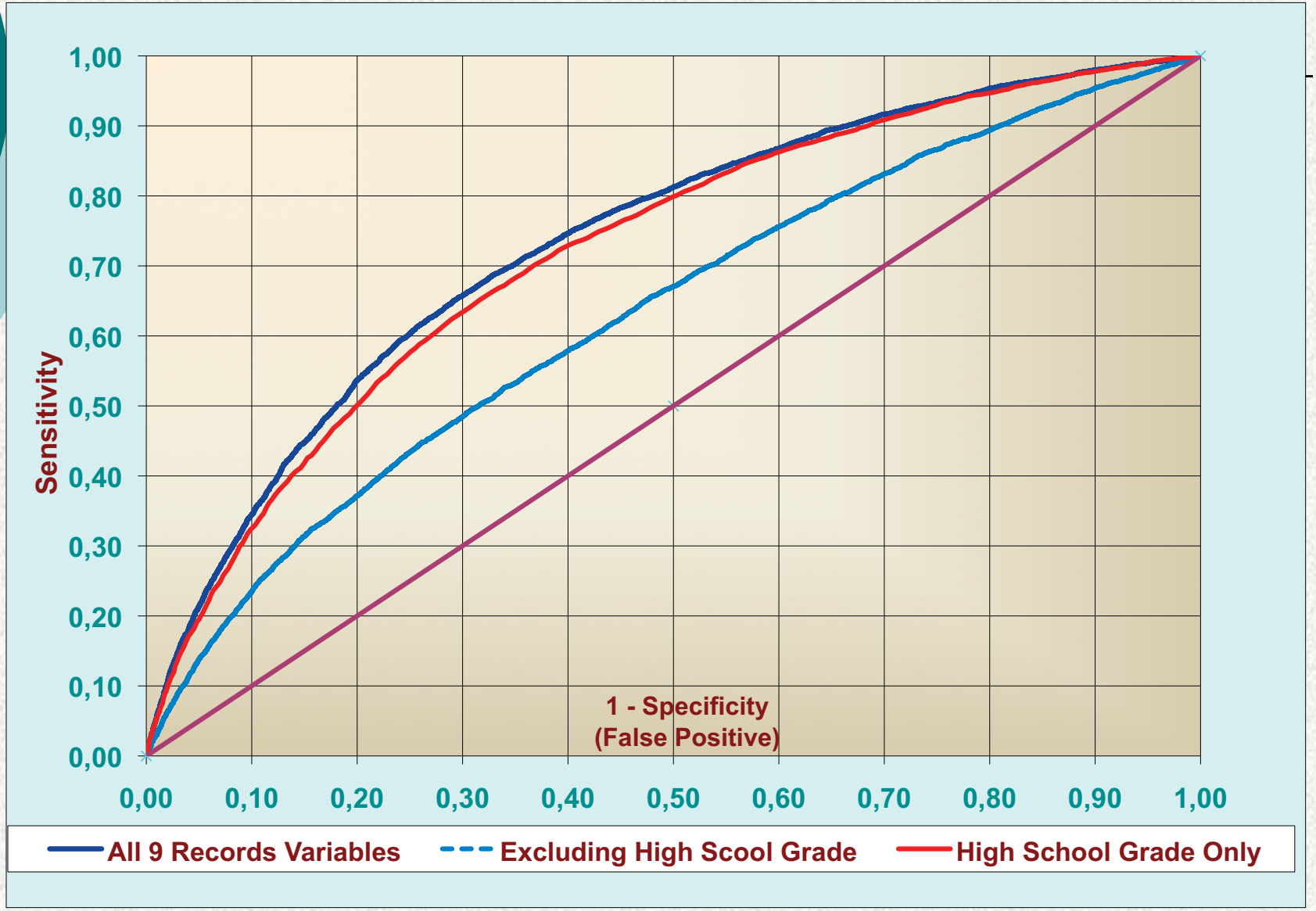
Null hypothesis
Area = 0.5

.90-1 = excellent (A)
.80-.90 = good (B)
.70-.80 = fair (C)
.60-.70 = poor (D)
.50-.60 = fail (F)



				95% Confidence Interval	
	Area	St Error	*Sig	Lower	Upper
F	0.687	0.023	0.00	0.641	0.733
M	0.766	0.025	0.00	0.716	0.816

‘Records’ Model (8 Variables)



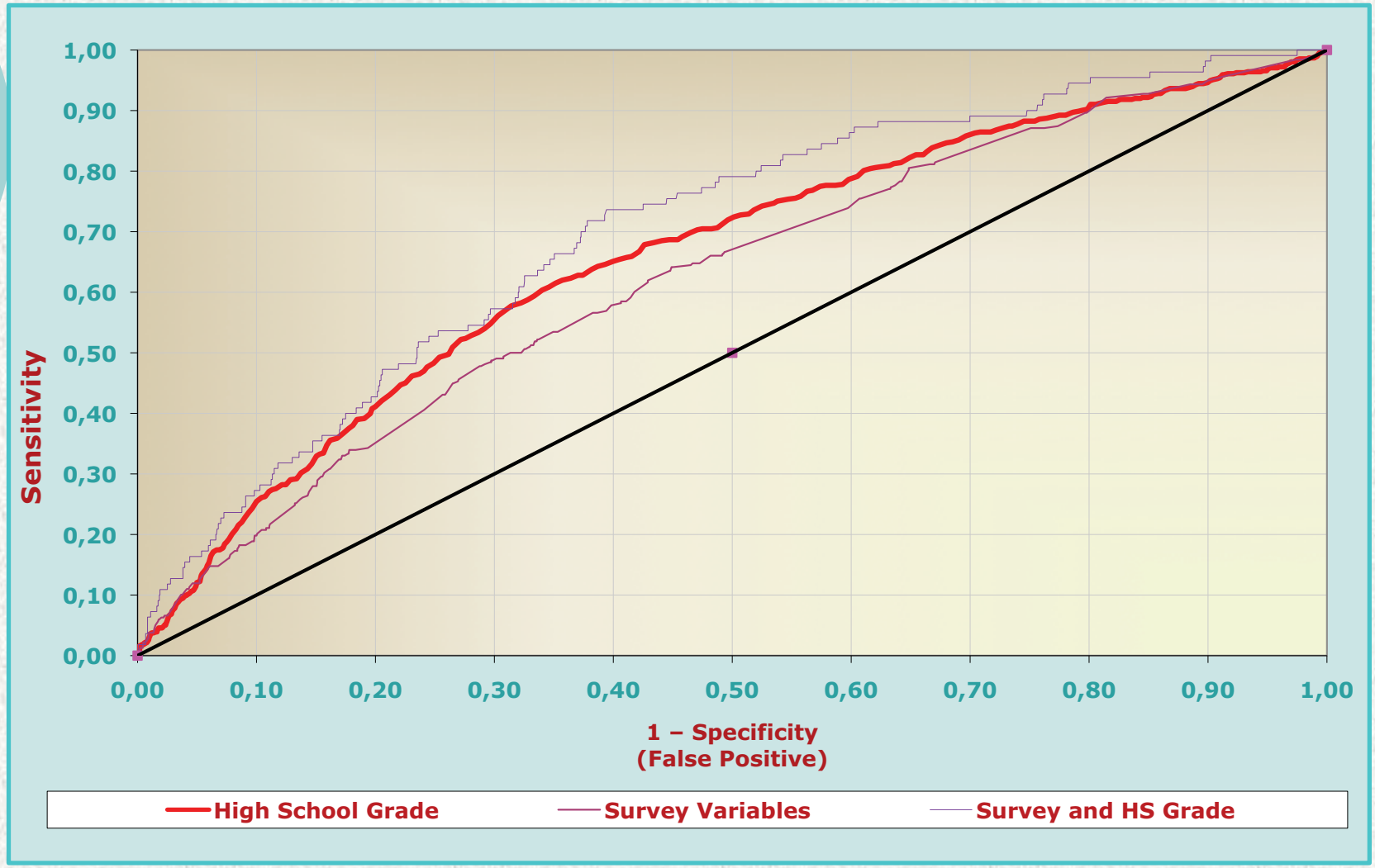
Accuracy of 'Records' Model - AUC

Test Result Variable(s)	Area under Curve (AUC)	Std. Error	Sig	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
High School Grade Alone (Poor)	**.659**	**0.012**	**0.000**	**0.636**	**0.683**
HS Grade + Records (8) (Poor)	.676	0.012	0.000	0.636	.686
8 Records Variables (Poor)	.608	0.012	0.000	0.585	0.631

Classification Matrix 'Records Model'

Cut-off = .16	N	Nagelkerke R^2	% Drop Out Correctly Classified (Sensitivity)	% Retained Correctly Classified (Specificity)	%Total Correctly Classified
HS Grade & 8 Survey variables	4153	.077	58.7%	69.5% (FP = 30.5%)	67.9%
8 Variables (Exclude SecV)	4427	.026	46.8%	70.3% (FP = 29.7%)	66.7%
High School Grade Only	4164	.063	59.4%	66.7% (FP =33.3%)	65.6%

Survey Model (9 Variables)



‘Survey’ Vs ‘Records’ Models

Test Results Variables	Area under Curve (AUC)	Std. Error	Sig	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
1. High School Grade (HSG) (Poor)	.659	.012	.000	.636	.683
2. HSG & 8 Records Variables (Poor)	.676	.012	.000	.636	.686
3. Records Variables(8) (Poor)	.608	.012	.000	.585	.631
4. Survey Variables (9) (Poor)	.625	.017	.000	.592	.658
5. HSG & 9 Survey Variables **(Fair)**	.700	.025	.000	.652	.749
6. All Variables (17) (Poor)	.672	.024	.000	.626	.718
7. HSG & All Variables **(Fair)**	.715	.025	.000	.665	.764

Variance Explained

	Model	Nagelkerke R^2
1	HS Grade	0.063
2	Records Variables (8)	0.026
3	Records (8) + HS Grade	0.077
4	Survey Variables (9)	0.044
5	Survey Variables (9) + HS Grade	0.089
6	Survey & Records (17 variables)	0.070
7	Survey (9) & Records (8) & HS Grade	0.104

Classification Accuracy

	Cutoff = .16				
	Model	**Sensitivity**	**Specificity**	**1-Specificity**	**Overall**
1	HS Grade	.594	.667	.333	.656
2	Records (8)	.468	.703	.297	.667
3	Records (8)+ HS Grade	.587	.695	.305	.679
4	Survey Variables (9)	.500	.687	.313	.659
5	Survey Variables (9) + HS Grade	.518	.723	.277	.695
6	Survey (9) & Records (8)	.514	.721	.279	.691
7	Survey (8) & Records (9) & SecV	.567	.742	.258	.718

Application – The Best Model?

Known:	
Historical attrition Rate to 3rd semester	16%
Historical retention rate to the 3rd semester	84%
For each cutoff and model:	
The model coefficients – calculate probabilities for each student	
The accuracy of classifying attrition (the percent of students who do drop out who are classified correctly by the model)	eg 57%
The false positive rate (% of retained students who are classified as dropping out)	eg 33%

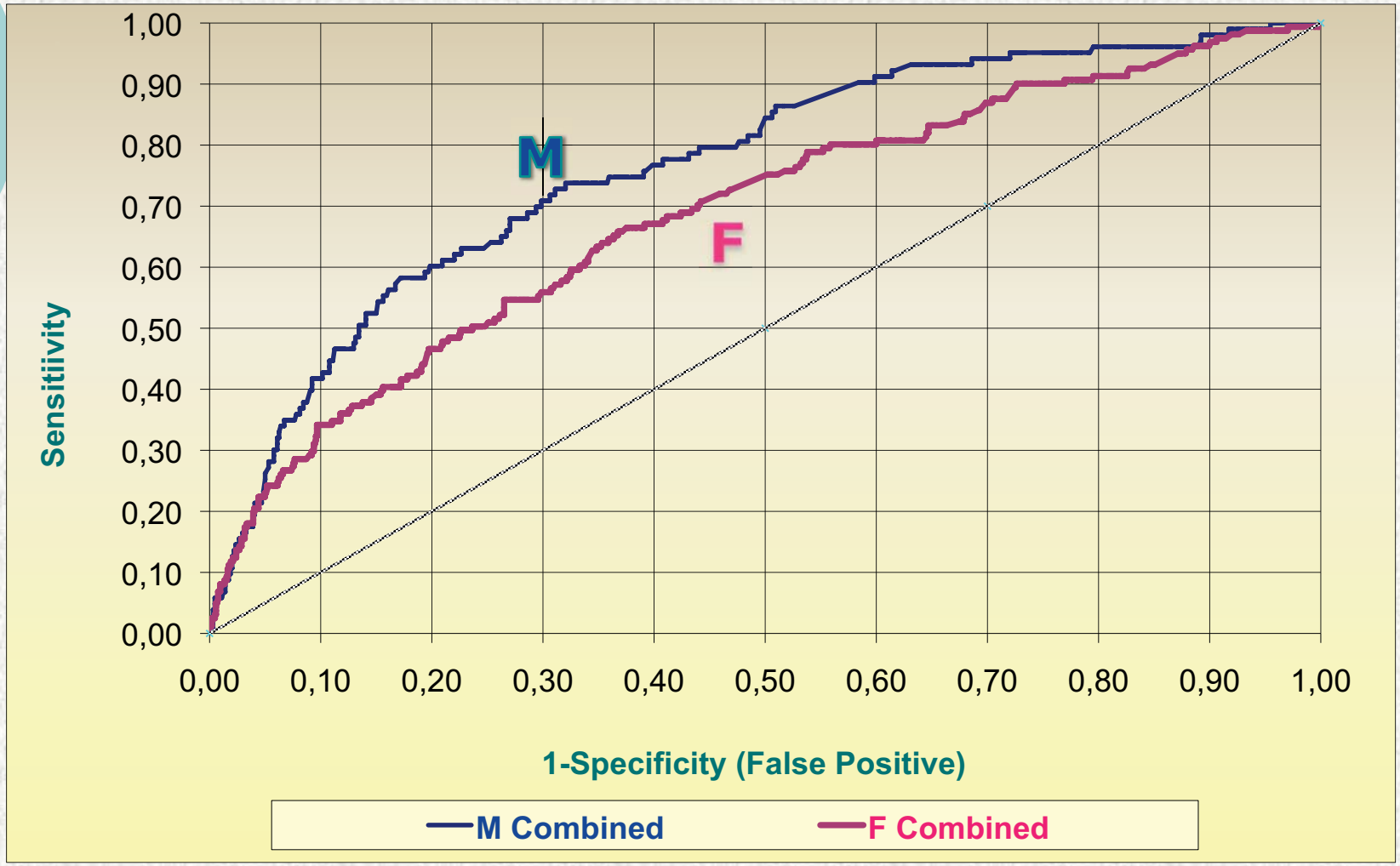
Application – 1000 New Students Cutoff .16

Mod	Historical (16% Att)	Classify	Model Predicted	Total Attrition Predicted	% Correct	
1	Att: 160 Ret: 840	Sens: .594 FP .333	95 280	375	25.4%	1:2.9
2	Att: 160 Ret: 840	Sens: .468 FP .297	75 249	324	23.1%	1:3.3
3	Att: 160 Ret: 840	Sens: .587 FP .305	94 256	350	26.8%	1:2.7
4	Att: 160 Ret: 840	Sens: .500 FP .313	80 263	343	23.3%	1:3.3
5	Att: 160 Ret: 840	Sens: .518 FP .277	83 233	316	26.3%	1:2.8
6	Att: 160 Ret: 840	Sens: .514 FP .279	82 234	317	26.0%	1:2.8
7	Att: 160 Ret: 840	Sens: .567 FP .258	91 217	307	29.5%	1:2.4

Application – 1000 New Students
70 students for remediation program

Mod Cut-off	Historical (16% Att)	Classify	Model Predicted	Total Attrition Predicted	% Correct	
1 (.25)	Att: 160 Ret: 840	Sens: .148 FP: .059	24 50	74	32.3%	1:2.1
2 (.21)	Att: 160 Ret: 840	Sens: .126 FP .059	20 50	70	28.9%	1:2.5
3 (.26)	Att: 160 Ret: 840	Sens: .162 FP .052	26 44	70	37.2%	1:1.7
4 (.24)	Att: 160 Ret: 840	Sens: .148 FP: .061	24 51	75	31.6%	1:2.2
5 (.28)	Att: 160 Ret: 840	Sens: .164 FP: .052	26 44	70	37.5%	1:1.7
6 (.28)	Att: 160 Ret: 840	Sens: .150 FP .055	24 46	70	34.2%	1:1.9
7 (.30)	Att: 160 Ret: 840	Sens: .174 FP .050	28 42	70	40.0%	1:1.5

Optimizing Attrition Models



Compare Male and Female Models of Attrition

.16	Characteristics	F	M	F+M
a	Sensitivity	49.1%	63.1%	0.567
b	Specificity	77.5%	76.1%	0.742
c	1 - Specificity (False Positives)	22.5%	23.9%	0.258
d	Nagelkerke R^2	0.105	0.195	0.104
e	Area Under ROC Curve	0.687	0.766	0.715
f	% New sample correct (Cutoff .16)	25.2%	29.8%	29.5%
g	Select 70 for remediation - % Correct	47.0%	49.1%	40.0%
h	Cutoff required for (g)	0.292	0.364	0.300

Summary

- Variability explained by all the models tested was low (Nagelkerke R^2)
- The accuracy of the models tested were judged to be poor to fair at best (Area under the ROC curve)
- Under certain conditions the HS grade and the more complete 'records' variables did as well or nearly as well as survey variables and high school grades
- Male and female models have different sensitivities at any given cutoff – and prediction can be improved by modeling the sexes separately

Summary

- The models tended to more accurately predict attrition for males than for females (Area under ROC curve, classification matrices)
- All models tested gave better than chance prediction
- None of the models predicted drop out particularly well
- The survey data used did improve the ability to predict attrition to a greater extent than the records variables in some situations, but not to the extent that we believe warrants the costs and overcomes the limitations of data collected through survey administration

Questions

	Differences in attrition rate between groups	Females	Males
Sig for both males and females	***Age – Was over 17 when starting college for the first time**	**17.9%**	**20.2%**
	***High school grade was < 75**	**16.0%**	**21.6%**
	Expected hours of paid employment was > 15 hours/ week	9.3%	12.5%
	Study Time <12 hours in last yr of study	6.7%	5.2%
	Motivation – Low or Average	6.5%	8.3%
	*Language was French	6.2%	3.7%
	*Median family income (post code) <$60000	4.9%	5.7%
	*English Placement Level - Low	2.9%	5.0%
	Place of birth father – in Canada	2.8%	4.5%
	*Diploma type - Technical	1.6%	3.9%
Sig for F only	**Student was not in first choice program**	**10.4%**	**2.9%**
	Anticipated study time at cegep	3.3%	1.8%
	*Country of birth – outside of Canada	2.7%	0.7%
Sig for M only	**Degree aspirations were DEC or Bachelor**	**3.0%**	**10.8%**
	Student was a first generation college student	1.4%	5.3%
Not Sig for either	Place of birth mother - Canada	1.8%	3.0%

Psychosocial and Study Skills Variables

(ACT Testing – Student Readiness Inventory)

- Academic discipline
- Academic self-confidence
- Commitment to college
- Communication skills
- Emotional control
- General determination
- Goal striving
- Social activity
- Social connection
- Study skills