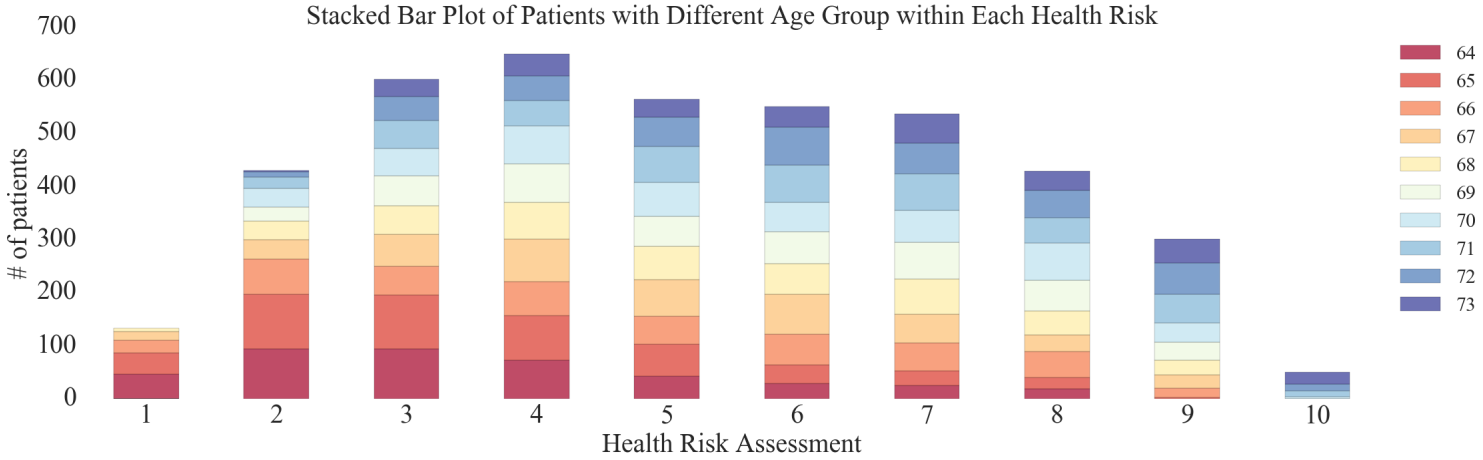
Identify Doctors' Effectiveness

- Doctor's effectiveness is defined by the likelihood that he/she will avoid patients' re-visit
- I explored patients population and failure% within each health risk, age, sex and doctor group
- I trained a logistic regression model
- I chose effective doctors and ineffective doctors based on model results (coefficients)

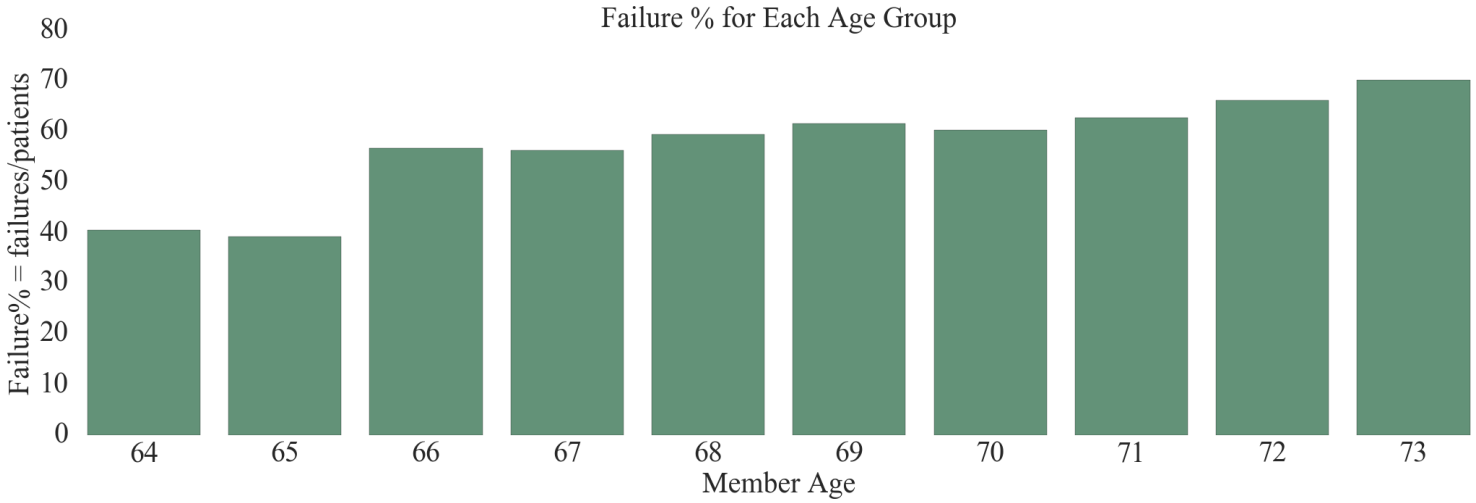
Patients Population in Different Health Risk Group

- The majority of patients with age 64-68 has health risk assessment from 1-5
- The majority of patients with age 69-70 has health riks assesment from 3-8
- Younger age group patients tend to have more healthy conditions



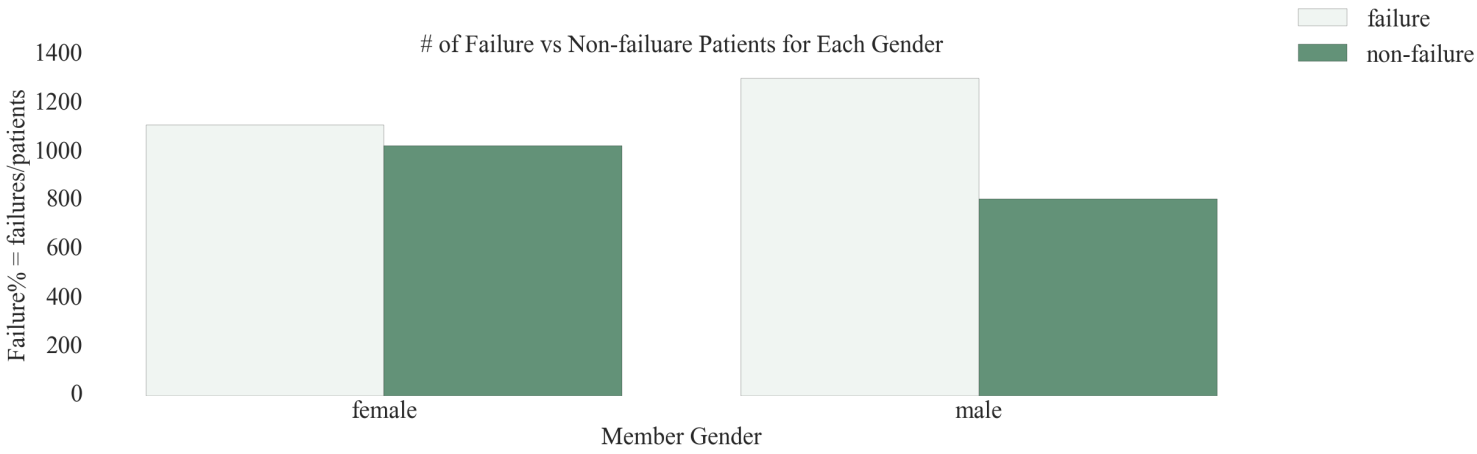
Failure Percentage within Each Age Group

- For patients with age 64 or 65 , the failure % is relatively low (around 40%)
- For patients with age from 66 to 70, the failure % is more or less the same (around 60%)
- For patients with age from 71 to 73, the failure % is slightly higher (around 65%)
- In general, patients in older age group are more likely to re-visit doctors



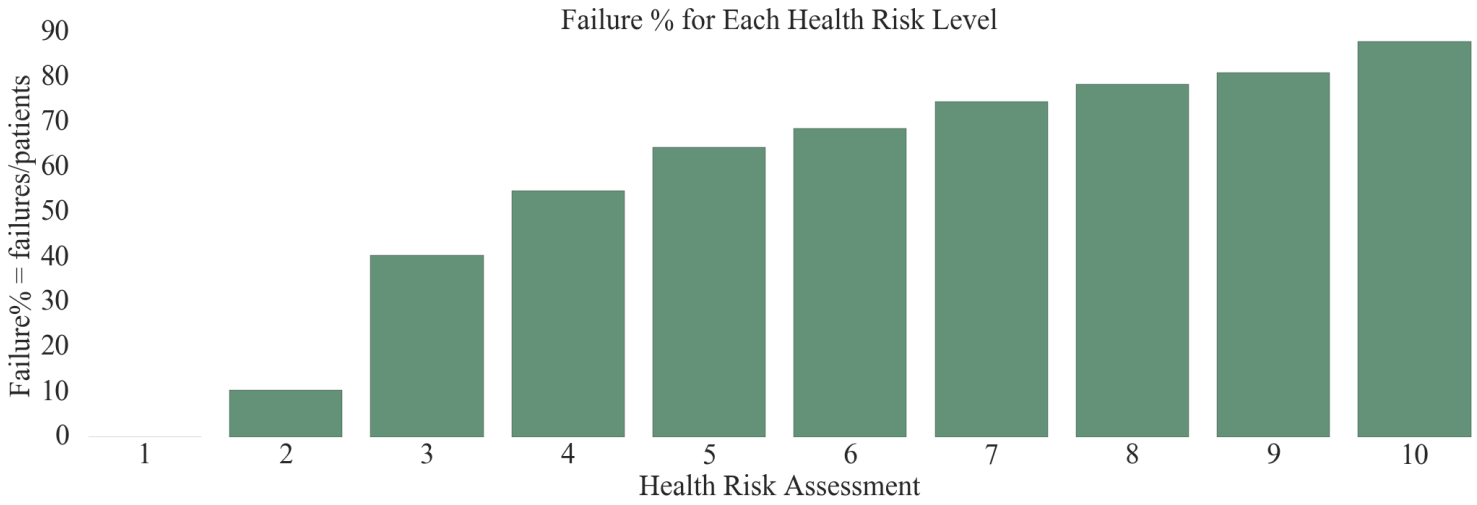
Failure Percentage for Each Gender

- For female patients, the number of patients revisited doctors is 50%
- For male patients, the number of patients re-visited doctors is 60%
- In general, male patients are more likely to re-visit doctors than female patients



Failure Percentage within Each Health Risk Assessment

- Failure percentage increases as health risk increases
- For patients with health risk = 10, almost 90% of them re-visited doctors
- Therefore, we should control for health risk assessment when evaluating doctors' effectiveness



Doctors' Failure Percentage

- The table shows the failure % for each doctor and the total number of patients they treat
- Only the top 20 doctors with lower failure % are shown
- The second table shows the proportion of high-risk patients a doctor treated, as you can see some doctors treated more sicker patients than others

	servicing_provider_name	failure_percentage	event_id
16	Demento	0.372093	43
62	Octopus	0.377778	45
74	Salk	0.404255	47
61	No	0.418605	43
33	Gray	0.425532	47
26	Faustus	0.431818	44
51	Lower	0.431818	44
43	Huntington	0.431818	44
21	Dre	0.463415	41
39	Horrible	0.479167	48
5	Beckett	0.489362	47
92	Xavier	0.490196	51
84	Toboggan	0.491803	61
30	Giggles	0.500000	42
88	Venture	0.500000	34
36	Hess	0.500000	40
69	Prentice	0.510204	49
48	Kelso	0.510638	47
11	Charcot	0.512195	41
85	Todd	0.513514	37

	servicing_provider_name	health_risk_assesment	Patient_Percentage
68	Broca	8	0.255814
736	Strangelove	8	0.244444
817	Who	8	0.239130
692	Seuss	9	0.200000
500	Martin	8	0.195122
454	Love	8	0.194444
492	Mario	9	0.194444
283	Gonzo	8	0.187500
555	No	8	0.186047
445	Lahiri	8	0.166667
699	Shephard	8	0.160714
481	Manhatten	8	0.157895
855	Zhivago	8	0.152174
32	Bashir	8	0.150000
319	Hawkeye	9	0.148936
399	Jeckyll	9	0.148148
195	Dre	9	0.146341
519	Mickhead	8	0.139535
232	Evil	9	0.138889
157	Dolittle	8	0.138889

Logistic Regression Modeling

- Assumed missing data in outcome column means patients are 'cured'
- The features Member Sex, Member Age and Health Risk Assessment are correlated
- Selected Health Risk Assessment as a continuous variable and Doctors as a categorical variable to build logistic regression
- Evaluated Doctor's effectiveness based on coefficients

	member_age	member_sex	health_risk_assesment
member_age	1.000000	-0.003566	0.380011
member_sex	-0.003566	1.000000	0.217608
health_risk_assesment	0.380011	0.217608	1.000000

One-Hot-Coding: servicing_provider_name is categorical variable.

```
Optimization terminated successfully.
         Current function value: 0.574941
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.572010
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.576973
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.576575
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.574257
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.575028
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.571224
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.574445
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.575411
         Iterations 6
Optimization terminated successfully.
         Current function value: 0.578583
         Iterations 6
```

68% accuracy on average

```
0.68472031076581574
```

fit the model using all data

```
Optimization terminated successfully.
         Current function value: 0.576216
         Iterations 6
```

- Dr.Andrews is chosen as my baseline
- The p-value of health risk assessment is significant
- The coefficient of health_risk_assesment is 0.45, which tells for one unit increase in health_risk_assessment, we will expect 50% increase in the Odds of failuare
- The p-values of all doctors are insignificant, which indicates that the doctors are not significantly better or worse than Dr.Andrews
- If a coefficient of a doctor is negative, it indiates that the outcome of failure is less likely, otherwise, it indicates that the outcome of failure is more likely

Logit Regression Results

Dep. Variable:	outcome	**No. Observations:**	4247
Model:	Logit	**Df Residuals:**	4150
Method:	MLE	**Df Model:**	96
Date:	Sun, 10 Apr 2016	**Pseudo R-squ.:**	0.1574
Time:	19:08:26	**Log-Likelihood:**	-2447.2
converged:	True	**LL-Null:**	-2904.2
		LLR p-value:	1.486e-133

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
Intercept	-2.0632	0.347	-5.951	0.000	-2.743 -1.384
health_risk_assesment	0.4528	0.018	25.624	0.000	0.418 0.487
Banting	0.5132	0.471	1.090	0.276	-0.409 1.436
Barnard	0.5121	0.482	1.062	0.288	-0.433 1.457
Bashir	-0.1205	0.490	-0.246	0.806	-1.081 0.840
Beardface	0.0588	0.482	0.122	0.903	-0.885 1.003
Beckett	-0.2839	0.464	-0.612	0.540	-1.193 0.625
Boom	0.5729	0.489	1.171	0.242	-0.386 1.532
Broca	-0.1704	0.482	-0.354	0.724	-1.115 0.774
Brown	-0.0344	0.460	-0.075	0.940	-0.936 0.867
Cameron	0.0035	0.470	0.008	0.994	-0.919 0.926
Castellano	-0.0753	0.481	-0.157	0.876	-1.018 0.868
Charcot	-0.0515	0.483	-0.107	0.915	-0.998 0.895
Constantine	0.2730	0.449	0.608	0.543	-0.608 1.154
Cox	0.7316	0.495	1.477	0.140	-0.240 1.703
Crane	0.1626	0.480	0.339	0.735	-0.777 1.102
Cuddy	0.3094	0.504	0.614	0.539	-0.678 1.297
Demento	-0.5779	0.491	-1.177	0.239	-1.540 0.384
Dolittle	0.5351	0.516	1.038	0.299	-0.475 1.546
Doom	0.2802	0.446	0.629	0.530	-0.594 1.154
Dorian	0.0857	0.468	0.183	0.855	-0.831 1.003
Doyle	0.2096	0.450	0.466	0.641	-0.673 1.092
Dre	-0.6453	0.491	-1.313	0.189	-1.609 0.318

Drew	0.5013	0.471	1.064	0.288	-0.423 1.425
Dudemeister	0.1302	0.478	0.272	0.785	-0.807 1.068
Ehrlich	0.1324	0.462	0.287	0.774	-0.773 1.038
Evil	0.2419	0.505	0.480	0.632	-0.747 1.231
Faustus	-0.6080	0.477	-1.276	0.202	-1.542 0.326
Feelgood	0.0200	0.463	0.043	0.966	-0.888 0.928
Fleming	0.1815	0.482	0.377	0.706	-0.763 1.126
Foreman	0.5341	0.473	1.130	0.259	-0.392 1.461
Giggles	-0.0269	0.472	-0.057	0.955	-0.953 0.899
Gonzo	0.1374	0.467	0.294	0.769	-0.778 1.053
Gooden	0.5041	0.510	0.988	0.323	-0.496 1.504
Gray	-0.3510	0.467	-0.752	0.452	-1.265 0.563
Haeckel	0.3929	0.484	0.812	0.417	-0.556 1.342
Hawkeye	-0.0057	0.473	-0.012	0.990	-0.933 0.922
Hess	-0.2802	0.495	-0.566	0.571	-1.250 0.690
Hibbert	0.1364	0.455	0.300	0.764	-0.755 1.028
Higgins	0.0645	0.475	0.136	0.892	-0.867 0.996
Horrible	-0.2587	0.464	-0.557	0.577	-1.169 0.651
House	0.3322	0.467	0.712	0.477	-0.583 1.247
Howser	0.0450	0.468	0.096	0.923	-0.872 0.962
Hunter	0.0990	0.481	0.206	0.837	-0.844 1.042
Huntington	-0.4649	0.475	-0.978	0.328	-1.396 0.466
Jeckyll	0.1529	0.459	0.333	0.739	-0.747 1.053
John	0.0304	0.488	0.062	0.950	-0.927 0.988
Jones	-0.0690	0.468	-0.147	0.883	-0.986 0.848
Katz	-0.0180	0.485	-0.037	0.970	-0.969 0.933
Kelso	-0.2826	0.467	-0.605	0.545	-1.198 0.633
Lahiri	-0.2079	0.507	-0.410	0.682	-1.202 0.787
Love	0.0787	0.514	0.153	0.878	-0.930 1.087
Lower	-0.6309	0.471	-1.339	0.181	-1.555 0.293
Mabuse	0.0290	0.467	0.062	0.950	-0.886 0.944
Manhatten	0.2450	0.455	0.539	0.590	-0.646 1.136
Mario	0.0233	0.510	0.046	0.964	-0.976 1.022

Martin	0.1666	0.488	0.342	0.733	-0.789 1.122
McCoy	-0.2898	0.462	-0.627	0.531	-1.196 0.616
Mickhead	0.1228	0.478	0.257	0.797	-0.814 1.060
Mindbender	0.1473	0.467	0.315	0.752	-0.768 1.063
Minot	0.3068	0.507	0.606	0.545	-0.686 1.300
Nick	0.3272	0.491	0.666	0.505	-0.635 1.290
No	-0.5642	0.487	-1.157	0.247	-1.520 0.391
Octopus	-0.7540	0.479	-1.572	0.116	-1.694 0.186
Orpheus	0.2501	0.478	0.523	0.601	-0.687 1.188
Oz	0.3160	0.469	0.674	0.500	-0.603 1.235
Penfield	0.2903	0.531	0.547	0.584	-0.750 1.330
Pepper	0.2743	0.482	0.569	0.569	-0.670 1.219
Perkins	0.3452	0.486	0.710	0.478	-0.608 1.298
Phil	0.5529	0.459	1.204	0.228	-0.347 1.453
Prentice	-0.0595	0.463	-0.129	0.898	-0.966 0.847
Quinn	0.0986	0.533	0.185	0.853	-0.947 1.144
Reed	-0.1894	0.553	-0.343	0.732	-1.274 0.895
Reid	0.2900	0.468	0.619	0.536	-0.628 1.208
Sacks	0.5699	0.496	1.149	0.251	-0.402 1.542
Salk	-0.4727	0.476	-0.993	0.321	-1.405 0.460
Scully	0.3495	0.466	0.750	0.453	-0.563 1.262
Seuss	0.2894	0.499	0.580	0.562	-0.689 1.268
Shephard	0.5301	0.458	1.157	0.247	-0.368 1.428
Snow	-0.0002	0.473	-0.001	1.000	-0.928 0.928
Spaceman	0.7902	0.490	1.611	0.107	-0.171 1.751
Spock	0.1301	0.498	0.261	0.794	-0.846 1.107
Strangelove	-0.3088	0.477	-0.647	0.518	-1.244 0.626
Tam	0.2369	0.472	0.502	0.616	-0.688 1.162
Teeth	0.4979	0.461	1.079	0.280	-0.406 1.402
Toboggan	-0.4691	0.439	-1.069	0.285	-1.329 0.391
Todd	-0.2174	0.492	-0.442	0.659	-1.182 0.747
Turk	-0.0503	0.466	-0.108	0.914	-0.964 0.863
Urbani	0.1955	0.505	0.387	0.698	-0.793 1.184

Venture	-0.1693	0.510	-0.332	0.740	-1.169 0.831
Watson	0.1668	0.469	0.356	0.722	-0.752 1.086
Who	-0.2087	0.476	-0.439	0.661	-1.141 0.723
Worm	0.0631	0.489	0.129	0.897	-0.896 1.022
Xavier	-0.0924	0.458	-0.202	0.840	-0.990 0.805
Zaius	-0.0732	0.483	-0.152	0.880	-1.020 0.874
Zhivago	0.4139	0.475	0.871	0.384	-0.517 1.345
de Chauliac	0.0884	0.476	0.185	0.853	-0.845 1.022

	Exp(Coef)
Spaceman	2.203920
Cox	2.078344
Boom	1.773314
Sacks	1.768040
Phil	1.738235
Dolittle	1.707630
Foreman	1.705943
Shephard	1.699168
Banting	1.670622
Barnard	1.668773
Gooden	1.655555
Drew	1.650919
Teeth	1.645209
health_risk_assesment	1.572774
Zhivago	1.512658
Haeckel	1.481253
Scully	1.418355
Perkins	1.412219
House	1.393979
Nick	1.387110
Oz	1.371624
Cuddy	1.362648
Minot	1.359014
Penfield	1.336832
Reid	1.336388
Seuss	1.335634
Doom	1.323447
Pepper	1.315636
Constantine	1.313925
Orpheus	1.284181
...	...

Turk	0.950960
Charcot	0.949767
Prentice	0.942218
Jones	0.933353
Zaius	0.929387
Castellano	0.927457
Xavier	0.911757
Bashir	0.886521
Venture	0.844264
Broca	0.843285
Reed	0.827422
Lahiri	0.812289
Who	0.811634
Todd	0.804606
Horrible	0.772020
Hess	0.755664
Kelso	0.753790
Beckett	0.752816
McCoy	0.748400
Strangelove	0.734355
Gray	0.703995
Huntington	0.628167
Toboggan	0.625592
Salk	0.623291
No	0.568803
Demento	0.561092
Faustus	0.544459
Lower	0.532120
Dre	0.524525
Octopus	0.470485

96 rows × 1 columns

- Based on the coefficents, I pick 3 doctors whom are considered as effective: Octopus, Dre, Lower
- I also pick 3 doctors whom are considered as ineffective: Spaceman, Cox and Boom
- The plot below shows the probability that a patient will re-visit a doctor if he/she is treated by one of the 6 doctors
- As you can see from the plot, for a patient with the same health risk assessment, the probability that the patient will re-vist is higher if he/she is treated by doctor Spaceman, Cox, or Booom