We're Number One! Does It Matter?

For the first time in Yale's history, they have been voted number one in a college hockey poll (two, actually). This pick is controversial for several reasons:

- They play in a conference regarded as, at best, middle-of-the road, the ECAC. This regard is partly based on interleague play, but is usually based on the poor performance of ECAC teams in the final NCAA Tournament in recent years. Indeed, based on that play alone, the ECAC looks much worse than its interleague schedule.

- As an Ivy League school, they are only permitted seven out-of-conference games, and only one of those games was scheduled with a team generally acknowledged to be at the best level of college hockey (Colorado College).

- Yale is not a traditional hockey power, and thus inherently deserving less respect.

This brief note is not intended to respond to any of these concerns. Instead, it takes as a fact that Yale was, despite these shortcomings, voted best in the country (in December). The questions I wish to look at is: what, if anything does this mean? In particular:

- What does being ranked first in December tell you about your ranking in March at the end of the season?

- What do rankings say anyway? Are higher ranked teams more likely to win games in the NCAA Tournament? Are there big differences between being ranked first and, say, fifth? And forget winning games. Are they more likely to win the tournament itself?

It is important to see that I have no idea what voters think they are supposed to be doing when they cast their votes. Thus, some voters may be explicitly casting their votes as to who has played well up to that point without any consideration of how well that team will do in the tournament. Others may be voting exclusively on how well they think the team will do in the tournament given the evidence up to that point. From my standpoint it doesn't matter why the voters did what they did: I just want to know what it means, if anything.

There are at least three metrics for the votes. First, we can use the ranking itself. This is slightly complicated by the fact that the rankings changed over the period from ranking 15 teams to ranking 20. Fortunately, all teams receiving votes are counted and we can generate a rank even if a team is not in the top 15 or 20. In the few cases where a team got no votes, I ranked them just below the lowest ranking of a team that got votes.

Second we can use the point scale used to generate the votes. Thus, in a 20 team field, a first place vote gets you twenty points and a second place vote gets you nineteen points. To make these vote totals comparable across years, though, one must make an adjustment for the maximum possible votes. I haven't done that analysis yet.

Finally, we can count first place votes. This measure at least tells you that *somebody* thought you were number one, and the more such votes you get the better.

I. The Data

I have gone back to 1993, the first year in which the NCAA Tournament was scheduled as a single elimination four round tournament. Before that period, high-ranked teams got byes. By itself, this does not make the earlier data unusable, but it would require more analysis to make sense of, since there is almost surely a major advantage in getting a bye, much bigger than the advantage of playing a low-ranked team in the first round. Thus, I have data on 120 games: there are 15 games every year and 8 years of data. For each game, I have the following data:

Conference, December Ranking, December Points and December First Place Votes of Winner

Conference, Final Ranking, Final Points and Final First Place Votes of Loser

With these data, we can at least get some idea of what the votes mean, if anything.

II. Will You Love Me In March like You Did In December?

Being ranked first in December is an excellent sign of a good ranking in March, but it is not a good harbinger for being ranked first in March. The following Table shows the Final Ranking of teams ranked number one in December:

year	team	frank
2003	North Dakota	12
2004	North Dakota	2
2005	Minnesota	7
2006	Wisconsin	2
2007	Minnesota	2
2008	Miami	2
2009	Notre Dame	2
2010	Miami	1

Given that the NCAA tournament so that teams ranked 1-4 are essentially in the same position (this abstracts from the difference between pairwise rankings and voting rankings) we see that being ranked first in December is certainly an excellent indications that you'll be highly ranked in March, and probably worthy of a number one seed in whatever regional you're in.

In general, December rankings are a pretty good indicator of March rankings. A simple linear regression yields:

frank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
drank	.5019368	.0603465	8.32	0.000	.3825129	.6213607
_cons	4.204774	.8320239	5.05	0.000	2.558223	5.851324

The way to read this is that best linear predictor of final rank is to take your December rank, halve it, and add 4. Thus, the best predictor if you're ranked 1st is about 4.7, while the best predictor if you're ranked 10th is about 9th. There are some technical reasons why this is not the best predictor for the upper and lower rankings, but I include it to show that there is a strong tendency for high and low rankings to persist. Note that this effect is biased upward somewhat, since you had to be good enough to make it into the tournament to be included in the regression.

OK. So it's no surprise that, unless they fall flat on their faces, Yale is very likely to make the tournament and, furthermore, to make it at a fairly high level. Ancillary evidene for this is available at the excellent site: http://siouxsports.com/hockey/rankings/rpidetails.php?teamid=31

Now we get to the interesting part. How well do December number ones do?

First, the data on December number ones over this period:

year	winner	loser	round	wdr	ldr	drankdif
2003	Ferris State	North Dakota	1	8	1	7
2004	North Dakota	Holy Cross	1	1	22	-21
2004	Denver	North Dakota	2	5	1	4
2005	Minnesota	Maine	1	1	15	-14
2005	Minnesota	Cornell	2	1	9	-8
2005	North Dakota	Minnesota	3	8	1	7
2006	Wisconsin	Bemidji State	1	1	19	-18
2006	Wisconsin	Cornell	2	1	8	-7
2006	Wisconsin	Maine	3	1	9	-8
2006	Wisconsin	BC	4	1	5	-4
2007	Minnesota	Air Force	1	1	31	-30
2007	North Dakota	Minnesota	2	17	1	16
2008	Miami	Air Force	1	1	22	-21
2008	BC	Miami	2	14	1	13
2009	Bemidji State	Notre Dame	1	22	1	21
2010	Miami	Alabama-Huntsville	1	1	25	-24
2010	Miami	Michigan	2	1	11	-10
2010	BC	Miami	3	3	1	2

December number ones have a record in the NCAA Tournament of 11-7. One of them, Wisconsin in 2006, won the Tournament outright. Of the losses, only one came to team much more lowly ranked in December: the 2007 second round loss of Minnesota to North Dakota, though in final rankings those teams were much closer – Minnesota was ranked second and North Dakota was ranked 6th. Twice, however, the December number one lost its first round match, in 2003 and 2009. Only three of the eight made the Frozen Four.

Here is the same table using Final rankings:

year	winner	loser	round	wfr	lfr	frankdif
2003	Cornell	Minnesota State	1	1	11	-10
2003	Cornell	BC	2	1	8	-7
2003	New Hampshire	Cornell	3	3	1	2
2004	Maine	Harvard	1	1	15	-14
2004	Maine	Wisconsin	2	1	11	-10
2004	Maine	BC	3	1	4	-3
2004	Denver	Maine	4	8	1	7
2005	Denver	Bemidji State	1	1	19	-18
2005	Denver	New Hampshire	2	1	6	-5
2005	Denver	Colorado College	3	1	3	-2
2005	Denver	North Dakota	4	1	10	-9
2006	Boston University	UNO	1	1	15	-14
2006	BC	Boston University	2	9	1	8
2007	Notre Dame	Alabama-Huntsville	1	1	25	-24
2007	Michigan State	Notre Dame	2	10	1	9
2008	Michigan	Niagara	1	1	19	-18
2008	Michigan	Clarkson	2	1	11	-10
2008	Notre Dame	Michigan	3	12	1	11
2009	Boston University	Ohio State	1	1	14	-13
2009	Boston University	New Hampshire	2	1	12	-11
2009	Boston University	Vermont	3	1	11	-10
2009	Boston University	Miami	4	1	13	-12
2010	Miami	Alabama-Huntsville	1	1	30	-29
2010	Miami	Michigan	2	1	23	-22
2010	BC	Miami	3	5	1	4

Using Final Rankings, the number ones are a collective 19-6. Again, there few huge upsets here: Notre Dame's 2008 loss to 12th ranked Michigan was in the Frozen Four, by which time Michigan would presumably been higher ranked if another vote had been taken. No final number one in this data lost its first round match, but BU in 2006 and Notre Dame in 2007 failed to make the Frozen Four. This time there are two outright winners: Denver and BU.

III. Last Question

There remains one other question: Do number ones do better because they are better, or because they have easier draws? To help answer this question, I have run a number of statistical models on the probability of a win. The best of these models yield the following results:

Difference	DRANK	FRANK
1	40.1%	48.5%
2	42.6%	50.3%
3	45.2%	52.0%
4	47.8%	53.8%
5	50.4%	55.5%
6	53.0%	57.2%
7	55.5%	58.9%
8	58.1%	60.6%
9	60.6%	62.3%
10	63.0%	63.9%
11	65.4%	65.5%
12	67.7%	67.0%
13	70.0%	68.6%
14	72.1%	70.1%
15	74.1%	71.5%
16	76.1%	72.9%
17	77.9%	74.3%
18	79.7%	75.6%
19	81.3%	76.8%
20	82.8%	78.1%

Broadly speaking, December Ranking differences, are, when you get to the Tournament , just about as accurate as Final Ranking differences. There is an anomaly at very small ranking differences in which your probably of winning a game doesn't exceed 50 percent until your December Ranking is 5 notches higher than your opponent, but that largely comes about from matches in which December rankings changed quite a bit. In any case, this model isn't really accurate enough to tell the difference between a 45 percent chance of winning and a 20 percent chance of winning. Also note that this model is based on all the data. If we restrict the model to teams which are number one, the results are noisier but similar.

IV. Conclusions?

Voting Yale number one suggests that they are a pretty good team, but a team guaranteed very little more than the average team in the tournament beyond a somewhat easier row to ho if they manage to maintain a high ranking.

We're Number One! Does It Matter?

For the first time in Yale's history, they have been voted number one in a college hockey poll (two, actually). This pick is controversial for several reasons:

- They play in a conference regarded as, at best, middle-of-the road, the ECAC. This regard is partly based on interleague play, but is usually based on the poor performance of ECAC teams in the final NCAA Tournament in recent years. Indeed, based on that play alone, the ECAC looks much worse than its interleague schedule.

- As an Ivy League school, they are only permitted seven out-of-conference games, and only one of those games was scheduled with a team generally acknowledged to be at the best level of college hockey (Colorado College).

- Yale is not a traditional hockey power, and thus inherently deserving less respect.

This brief note is not intended to respond to any of these concerns. Instead, it takes as a fact that Yale was, despite these shortcomings, voted best in the country (in December). The questions I wish to look at is: what, if anything does this mean? In particular:

- What does being ranked first in December tell you about your ranking in March at the end of the season?

- What do rankings say anyway? Are higher ranked teams more likely to win games in the NCAA Tournament? Are there big differences between being ranked first and, say, fifth? And forget winning games. Are they more likely to win the tournament itself?

It is important to see that I have no idea what voters think they are supposed to be doing when they cast their votes. Thus, some voters may be explicitly casting their votes as to who has played well up to that point without any consideration of how well that team will do in the tournament. Others may be voting exclusively on how well they think the team will do in the tournament given the evidence up to that point. From my standpoint it doesn't matter why the voters did what they did: I just want to know what it means, if anything.

There are at least three metrics for the votes. First, we can use the ranking itself. This is slightly complicated by the fact that the rankings changed over the period from ranking 15 teams to ranking 20. Fortunately, all teams receiving votes are counted and we can generate a rank even if a team is not in the top 15 or 20. In the few cases where a team got no votes, I ranked them just below the lowest ranking of a team that got votes.

Second we can use the point scale used to generate the votes. Thus, in a 20 team field, a first place vote gets you twenty points and a second place vote gets you nineteen points. To make these vote totals comparable across years, though, one must make an adjustment for the maximum possible votes. I haven't done that analysis yet.

Finally, we can count first place votes. This measure at least tells you that *somebody* thought you were number one, and the more such votes you get the better.

I. The Data

I have gone back to 1993, the first year in which the NCAA Tournament was scheduled as a single elimination four round tournament. Before that period, high-ranked teams got byes. By itself, this does not make the earlier data unusable, but it would require more analysis to make sense of, since there is almost surely a major advantage in getting a bye, much bigger than the advantage of playing a low-ranked team in the first round. Thus, I have data on 120 games: there are 15 games every year and 8 years of data. For each game, I have the following data:

Conference, December Ranking, December Points and December First Place Votes of Winner

Conference, Final Ranking, Final Points and Final First Place Votes of Loser

With these data, we can at least get some idea of what the votes mean, if anything.

II. Will You Love Me In March like You Did In December?

Being ranked first in December is an excellent sign of a good ranking in March, but it is not a good harbinger for being ranked first in March. The following Table shows the Final Ranking of teams ranked number one in December:

year	team	frank
2003	North Dakota	12
2004	North Dakota	2
2005	Minnesota	7
2006	Wisconsin	2
2007	Minnesota	2
2008	Miami	2
2009	Notre Dame	2
2010	Miami	1

Given that the NCAA tournament so that teams ranked 1-4 are essentially in the same position (this abstracts from the difference between pairwise rankings and voting rankings) we see that being ranked first in December is certainly an excellent indications that you'll be highly ranked in March, and probably worthy of a number one seed in whatever regional you're in.

In general, December rankings are a pretty good indicator of March rankings. A simple linear regression yields:

frank	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
drank	.5019368	.0603465	8.32	0.000	.3825129	.6213607
_cons	4.204774	.8320239	5.05	0.000	2.558223	5.851324

The way to read this is that best linear predictor of final rank is to take your December rank, halve it, and add 4. Thus, the best predictor if you're ranked 1st is about 4.7, while the best predictor if you're ranked 10th is about 9th. There are some technical reasons why this is not the best predictor for the upper and lower rankings, but I include it to show that there is a strong tendency for high and low rankings to persist. Note that this effect is biased upward somewhat, since you had to be good enough to make it into the tournament to be included in the regression.

OK. So it's no surprise that, unless they fall flat on their faces, Yale is very likely to make the tournament and, furthermore, to make it at a fairly high level. Ancillary evidence for this is available at the excellent site: http://siouxsports.com/hockey/rankings/rpidetails.php?teamid=31

Now we get to the interesting part. How well do December number ones do?

First, the data on December number ones over this period:

year	winner	loser	round	wdr	ldr	drankdif
2003	Ferris State	North Dakota	1	8	1	7
2004	North Dakota	Holy Cross	1	1	22	-21
2004	Denver	North Dakota	2	5	1	4
2005	Minnesota	Maine	1	1	15	-14
2005	Minnesota	Cornell	2	1	9	-8
2005	North Dakota	Minnesota	3	8	1	7
2006	Wisconsin	Bemidji State	1	1	19	-18
2006	Wisconsin	Cornell	2	1	8	-7
2006	Wisconsin	Maine	3	1	9	-8
2006	Wisconsin	BC	4	1	5	-4
2007	Minnesota	Air Force	1	1	31	-30
2007	North Dakota	Minnesota	2	17	1	16
2008	Miami	Air Force	1	1	22	-21
2008	BC	Miami	2	14	1	13
2009	Bemidji State	Notre Dame	1	22	1	21
2010	Miami	Alabama-Huntsville	1	1	25	-24
2010	Miami	Michigan	2	1	11	-10
2010	BC	Miami	3	3	1	2

December number ones have a record in the NCAA Tournament of 11-7. One of them, Wisconsin in 2006, won the Tournament outright. Of the losses, only one came to team much more lowly ranked in December: the 2007 second round loss of Minnesota to North Dakota, though in final rankings those teams were much closer – Minnesota was ranked second and North Dakota was ranked 6th. Twice, however, the December number one lost its first round match, in 2003 and 2009. Only three of the eight made the Frozen Four.

Here is the same table using Final rankings:

year	winner	loser	round	wfr	lfr	frankdif
2003	Cornell	Minnesota State	1	1	11	-10
2003	Cornell	BC	2	1	8	-7
2003	New Hampshire	Cornell	3	3	1	2
2004	Maine	Harvard	1	1	15	-14
2004	Maine	Wisconsin	2	1	11	-10
2004	Maine	BC	3	1	4	-3
2004	Denver	Maine	4	8	1	7
2005	Denver	Bemidji State	1	1	19	-18
2005	Denver	New Hampshire	2	1	6	-5
2005	Denver	Colorado College	3	1	3	-2
2005	Denver	North Dakota	4	1	10	-9
2006	Boston University	UNO	1	1	15	-14
2006	BC	Boston University	2	9	1	8
2007	Notre Dame	Alabama-Huntsville	1	1	25	-24
2007	Michigan State	Notre Dame	2	10	1	9
2008	Michigan	Niagara	1	1	19	-18
2008	Michigan	Clarkson	2	1	11	-10
2008	Notre Dame	Michigan	3	12	1	11
2009	Boston University	Ohio State	1	1	14	-13
2009	Boston University	New Hampshire	2	1	12	-11
2009	Boston University	Vermont	3	1	11	-10
2009	Boston University	Miami	4	1	13	-12
2010	Miami	Alabama-Huntsville	1	1	30	-29
2010	Miami	Michigan	2	1	23	-22
2010	BC	Miami	3	5	1	4

Using Final Rankings, the number ones are a collective 19-6. Again, there are few huge upsets here: Notre Dame's 2008 loss to 12th ranked Michigan was in the Frozen Four, by which time Michigan would presumably been higher ranked if another vote had been taken. No final number one in this data lost its first round match, but BU in 2006 and Notre Dame in 2007 failed to make the Frozen Four. This time there are two outright winners: Denver and BU.

III. Last Question

There remains one other question: Do number ones do better because they are better, or because they have easier draws? To help answer this question, I have run a number of statistical models on the probability of a win. The best of these models yield the following results:

Difference	DRANK	FRANK
1	40.1%	48.5%
2	42.6%	50.3%
3	45.2%	52.0%
4	47.8%	53.8%
5	50.4%	55.5%
6	53.0%	57.2%
7	55.5%	58.9%
8	58.1%	60.6%
9	60.6%	62.3%
10	63.0%	63.9%
11	65.4%	65.5%
12	67.7%	67.0%
13	70.0%	68.6%
14	72.1%	70.1%
15	74.1%	71.5%
16	76.1%	72.9%
17	77.9%	74.3%
18	79.7%	75.6%
19	81.3%	76.8%
20	82.8%	78.1%

Broadly speaking, December Ranking differences, are, when you get to the Tournament , just about as accurate as Final Ranking differences. There is an anomaly at very small ranking differences in which your probably of winning a game doesn't exceed 50 percent until your December Ranking is 5 notches higher than your opponent, but that largely comes about from matches in which December rankings changed quite a bit. In any case, this model isn't really accurate enough to tell the difference between a 45 percent chance of winning and a 50 percent chance of winning. Also note that this model is based on all the data. If we restrict the model to teams which are number one, the results are noisier but similar.

IV. Conclusions?

Voting Yale number one suggests that they are a pretty good team, but a team guaranteed very little more than the average team in the tournament beyond a somewhat easier row to ho if they manage to maintain a high ranking.