

# Calculating Strength of Schedule, and Choosing Teams for March Madness

Paul Fearnhead and Benjamin M Taylor

*Department of Mathematics and Statistics, Lancaster University, UK*

## **Abstract**

We propose a new way of quantifying a team's strength of schedule for NCAA basketball. This strength of a schedule is defined as the number of games a team on the borderline of the annual national tournament would expect to win if they played that schedule. This gives a direct way of quantifying how well different teams have done relative to the schedules they have played. Our motivation for constructing this strength of schedule is to help inform the choice of teams given bids to the national tournament: teams who have won more games than their strength of schedule have strongest evidence that they deserve such a bid. Estimating the strength of schedules is possible through fitting a simple statistical model to the results of all regular season matches. We are able to quantify the uncertainty in these estimates, which helps differentiate between teams with clear evidence for selection and those on the borderline. We apply our method to data from the 2007/08 and 2008/09 season. Our results suggest that St. Mary's warranted a bid to the 2009 tournament, at the expense of Wisconsin; and that

both Arizona State and Nebraska warranted bids to the 2008 tournament instead of South Alabama and St. Joseph's.

*Keywords:* NCAA Basketball, RPI, Strength of Schedule,

## 1 Introduction

The highlight of the NCAA men's basketball season is *March Madness*, the national division I tournament held each March. The tournament has a field of 65 teams. After an initial play-in game, to reduce the field to 64 teams, there is a standard knock-out format. See Carlin (1996) for further details. The importance of the the annual tournament is highlighted by CBS paying \$6 billion for 11 years of TV rights in 1999.

The initial field of 65 teams consists of 31 conference champions, and 34 teams that are selected by an NCAA committee. These 34 teams are said to be given *at-large bids* to the tournament, The decision by this committee is announced on the Sunday, known as selection Sunday, a week prior to the start of the tournament.

The choice of the 34 teams to be given at-large bids is often contentious, with much media coverage prior to the announcement on selection Sunday (see the introduction of Harville, 2003, for more discussion). The difficulty in choosing who receives the at-large bids comes from the large differences in schedules, and the corresponding strength of opponents, of the different teams in contention. This means that it is not possible to purely compare and order teams based on their win-loss record. For example in 2009 Arizona, with a win-loss record of 19 – 13 were given a bid, whereas Creighton were not, despite a win-loss record

of 26 – 7. Our results in Section 4 support both these decisions, due to the substantially more difficult schedule that Arizona played.

We shall call the strength of a team’s opponents their *strength of schedule*. The aim of this paper is to construct a meaningful measure of a team’s strength of schedule that will make it possible to directly compare different teams’ win-loss records after accounting for the different opponents that they have played. Our strength of schedule is calculated by first considering a fictitious team of a certain standard. We then estimate on average how many games that team would win if they played a specific schedule. This expected number of wins is the strength of that schedule, and is a benchmark against which we can compare a team’s actual win-loss record. By choosing the standard of this fictitious team to be that of a team who is right on the borderline for receiving an at-large bid, we immediately get an indication of which teams have performed well enough to receive an at-large bid: they will have won more games than their strength of schedule indicates. Thus we claim this gives a simple, and easily interpretable, indication that could guide the decision made on selection Sunday. Calculation of each team’s strength of schedule is possible through applying a statistical model (Stefani, 1980; Harville and Smith, 1994) to the results of matches in the current season.

The outline of the paper is as follows. We first discuss existing statistics that are calculated for NCAA basketball teams, and outline the difficulties in using these to guide selection of teams for at-large bids. We then introduce a simple statistical model for analysing results of basketball matches, and show how this model can be used to estimate the probability that a team would win a given match. In Section 3 we then show how we can use this model, fitted to the current season’s results, to calculate each team’s strength of schedule. We will

also explain why we believe calculating our strength of schedule is more appropriate than using other inferences from the fitted model to help guide selection of teams for at-large bids. We then analyse data from both the 2007/08 and 2008/09 seasons in Section 4. These suggest that St. Mary's should have received a bid to the 2009 tournament instead of Wisconsin, and that Arizona State and Nebraska deserved bids to the 2008 tournament in place of South Alabama and St. Joseph's. The paper ends with a discussion.

## 2 Current use of statistics

Currently a number of statistical summaries of each team's performance are calculated to help guide which teams have performed best in a given season, and thus deserve an at-large bid. One system used to rank teams is the ranking percentage index (RPI). Initially it was calculated as a weighted average of team's winning percentage, their opponents' winning percentages, and their opponents' opponents' winning percentages. The weights are  $1/4$ ,  $1/2$  and  $1/4$  respectively. More recently it has been changed slightly to give more weight in a team's winning percentage to away wins and home losses, and less weight to home wins and away losses.

The RPI attempts to take account of both a team's winning percentage and its strength of schedule, although it does this in a somewhat *ad-hoc* manner. There have been a number of articles criticising the RPI (see e.g. cheap ticket to the dance: systematic bias in college basketball's rating percentage index, 2007, and references therein).

The RPI is often used to construct further summaries of a team's performance. This includes reporting a team's win-loss record against opponents who are

ranked in, say, the top 50 teams based on RPI. Even if the RPI was accurately defining the top 50 teams, there are problems with such summaries. The main problem is that it over-emphasises wins against teams just inside the top 50 as opposed to those against teams just outside; and by comparison it ignores the widely differing standards of teams within the top 50. These drawbacks are discussed in more detail in Section 4.

Related to the RPI, is a measure of a team's strength of schedule based on the RPI excluding the team's own winning percentage. This strength of schedule is often reported in terms of its rank relative to all other teams: so it gives information about which team has had the hardest schedule, the second hardest schedule, and so on.

We will highlight two problems with this approach to calculating and reporting teams' strength of schedule, both of which will be addressed by our approach to calculating a team's strength of schedule. The first problem is that the ranking of teams does not allow a simple and direct comparison of different team's records. For example which record was better in the 2008/09 season: Wisconsin's win-loss record of 19-13 with the 42nd hardest schedule, or St. Mary's one of 24-6 with the 118th hardest schedule? Answering this question is not possible without a way of quantifying, in terms of wins/losses, how much harder the first schedule is than the second.

The second problem is that the relative strength of different schedules will depend on the ability of the team that is being considered. For example consider two hypothetical schedules in 2008/09 each of two games, all at neutral venues: the first involves playing the University of North Carolina and Alcorn State (one of the strongest and one of the weakest teams); the other consists of playing Rutgers (a team ranked in the 100s) twice. For a top 25 team the latter would

be easier, with the team likely to win all matches except against North Carolina. Whereas for one of the weakest teams the former would be, with the team likely to lose all matches except against Alcorn State. Thus if we are calculating strength of schedules to help with the selection of teams for at-large bids, we need to take account of the ability of those teams that will be in contention.

### **3 A New Strength of Schedule**

Our approach addresses these two problems by calculating a team's strength of schedule in terms of the expected number of wins a team on the borderline of receiving an at-large bid would get if they played that schedule. We now describe how we can calculate this. We first present a standard statistical model for basketball data which we will use, and then show how after fitting this model we can calculate each team's strength of schedule.

#### **3.1 Modelling and Predicting Basketball Results**

Currently the NCAA men's basketball 1st division has 347 teams, with the number,  $n$ , of regular season matches (prior to selection Sunday) being around 5,000. Following Stefani (1977, 1980); Harville and Smith (1994); Harville (2003), we consider a linear model for the difference in the scores between the two teams in a basketball match. We assume the teams are numbered, arbitrarily, from 1 to 347. The model has parameters  $\beta_j$ ,  $j = 1, \dots, 347$  corresponding to the relative ability of the teams, and  $\gamma$  which corresponds to home advantage.

For match  $i$  let the teams be  $a_i$  and  $b_i$ , and let  $z_i$  take the value 1 if team  $a_i$  is at

home, -1 if team  $b_i$  is at home, and 0 if the match is at a neutral venue. Finally, let  $y_i$  be team  $a_i$ 's score less team  $b_i$ 's score at the end of regulation time. So a positive value corresponds to team  $a_i$  winning, and a negative score corresponds to team  $b_i$  winning, and 0 corresponds to the match entering overtime. Then our model is, for  $i = 1, \dots, n$ ,

$$y_i = \beta_{a_i} - \beta_{b_i} + \gamma z_i + \sigma \epsilon_i, \quad (1)$$

where  $\sigma$  is the standard error of the residuals, and we assume that each  $\epsilon_i$  is a realisation of an independent standard normal random variable.

It is not possible to identify all  $\beta$ s in this model, as adding or subtracting a constant to all the  $\beta$ s leaves the model unchanged. The interpretation of the  $\beta$ s is in terms of relative strength of teams: so if  $\beta_j > \beta_k$ , then  $\beta_j - \beta_k$  is the average number of points we would expect team  $j$  to beat team  $k$  by if they played at a neutral venue. The parameter  $\gamma$  is the average number of points advantage a team gets by playing at home. Note that we assume a normal distribution for the errors, whereas  $y_i$  must take integer values. A qq-plot of residuals of the fitted model to the 2008/09 data suggests that this approximation is reasonable, see Figure 1(a). Also, see Harville (2003) for more discussion on the use of a normal distribution for the errors in the linear model.

In constructing our strength of schedule in Section 3.2 we require to make predictions from our model. Thus to more naturally incorporate uncertainty in the  $\beta$  parameter estimates we use a Bayesian analysis. (There is little uncertainty in the estimates of  $\gamma$  and  $\sigma$  so we estimate these and treat them as fixed when making predictions.) For this we introduce priors for the  $\beta$  parameters. Our choice of prior takes account of information on which conference a team belongs to, and allows for differences in the average strength of each conference. There are 31 conferences, and we group together all

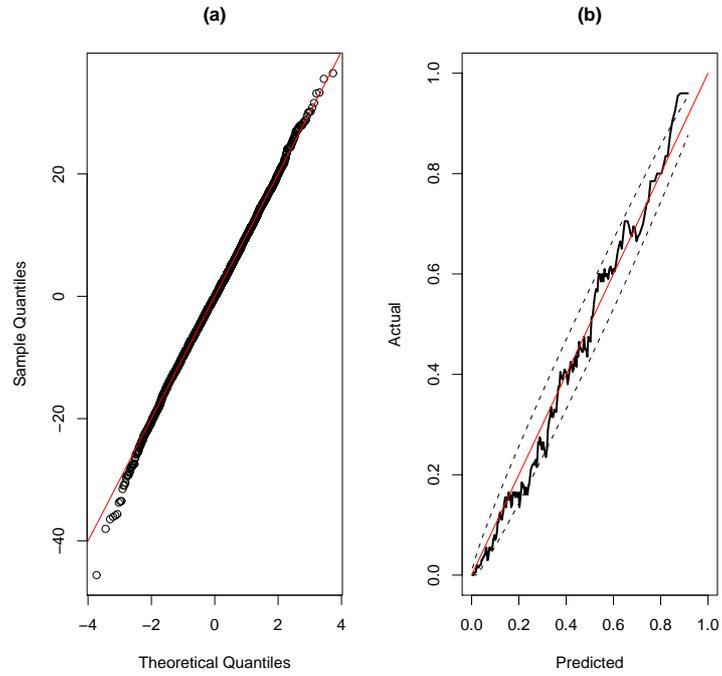


Figure 1: Plots of model-fit: (a) Quantile-quantile plot of residuals for the linear model against a standard normal distribution. (b) Plot of actual against predicted proportion of wins for the first named team across different sets of matches. All matches were ranked based on predicted probability of a win, and each set contains 400 consecutive matches from this ranked list. Dashed lines give approximate 95% confidence intervals calculated from the predicted probabilities.

non-conference teams. So we introduce mean conference strengths  $\eta_k$ , for  $k = 1, \dots, 32$ . We assume that if team  $j$  belongs to conference  $c_j$  then  $\beta_j$  has a normal prior with mean  $\eta_k$  and variance  $\tau_t^2$ ; and we assume normal priors for  $\eta_{c_j}$ s with mean 0 and variance  $\tau_c^2$ . The remaining parameters in the model,  $\gamma$ ,  $\sigma$ ,  $\tau_c$  and  $\tau_t$  are estimated by maximising the marginal likelihood.

One way of interpreting the priors is that they perform shrinkage. Note that the priors include no a priori information about which teams or which conferences tend to be stronger. Allowing for the differences in conference strengths is important because the majority of each team's matches are within conference, and without this structure the priors would over-shrink teams in good, or bad, conferences towards 0. The effect of the prior is noticeable when estimating parameters mid-way through the season, though has only a minor effect on estimates at the end of the season. By introducing this prior we avoid any problems of unidentifiability, and the effect of the prior is to center the estimates of the  $\beta$ s around an average of 0.

Previous analyses using this model (e.g. Harville, 2003) have used the resulting estimates of the  $\beta$ s to rank teams. However we argue that using such a ranking is not appropriate for choosing which teams should be given at-large bids. The reason for this is that the estimates of a  $\beta_j$  will depend only on the opponents of team  $j$  and team  $j$ 's net points difference in those matches. It does not directly depend on the number of matches that team  $j$  wins. However, we would want the choice of teams for the national tournament to be governed by their win-loss record (after accounting for their opponents). This is natural, as if two teams played identical schedules you would rank the team who won more matches higher, and not the team with the better net points difference. Also it means that a team's best strategy in each game is always to try and win the game, such

as by deliberately fouling to get the ball back towards the end of a game, even if this risks losing by a bigger margin.

### 3.2 Calculating the Strength of Schedule

Rather than using the estimates of the linear model (1) directly to rank teams, we instead use them to construct a strength of schedule for each team. This strength of schedule is based on calculating the expected number of matches a fictitious team, of fixed ability, would win if they played the schedule. Let the strength of this fictitious team be  $\beta_0$ . In order to calculate these strength of schedules, we need a way of estimating the probability a team of this strength would have of a winning any specific match.

Consider a match against team  $j$ , and let  $z$  be an indicator of which team is at home, defined as above. The linear model (1) gives that the probability that the fictitious team would win is

$$\Phi\left(\frac{\beta_0 - \beta_j + \gamma z}{\sigma}\right), \quad (2)$$

where  $\Phi$  is the distribution function of the standard normal distribution. This follows by considering the possible values of the error term in the linear model which corresponds to the fictitious team scoring more points than team  $j$ . Note that this calculation ignores the discrete nature of the difference in scores, and thus the possibility of the match entering over-time. However alternative approaches, for example based on over-time corresponding to an absolute difference in the two teams' scores being less than  $1/2$  and a subsequent model for the likelihood of each time winning in over-time, give negligible difference from using (2).

Empirically this approach to predicting the probability of wins is supported by

the normal qq-plot of the residuals of the linear model, see Figure 1(a). Also, for the actual regular season matches we compared predicted win probabilities of the first named team in each match, against actual frequency of wins. We calculated both mean predicted probability and frequency for sets of 400 matches. The sets of matches were obtained by first ranking the matches in terms of the predicted probability of a win, and then each set contains 400 consecutive matches from this list. This ensures that each set contains matches with similar predicted win probabilities. A plot of actual against predicted is shown in Figure 1(b), and shows the actual frequency to almost always lie within approximate 95% confidence regions obtained from the predicted probabilities.

We cannot directly use (2) as it depends on unknown parameters  $\beta_j$ ,  $\gamma$  and  $\sigma$ . Our approach is to plug in our estimates of  $\gamma$  and  $\sigma$ , and then average over the posterior distribution for  $\beta_j$ , given the data from all regular season matches. Note that the posterior for  $\beta_j$  can be calculated analytically, and the expectation over the posterior can be done numerically.

To calculate a teams strength of schedule, for a given  $\beta_0$ , we just sum these expected win probabilities for each of the matches that that team played. We can also estimate the posterior variance of the probability (2) for each match, and this can be used to determine an approximate standard error for the estimate of a team's strength of schedule.

As described above, our strength of schedule for each team will depend on  $\beta_0$ . The relative value of each team's strength of schedule will depend on this choice. For example in the limit as  $\beta_0 \rightarrow -\infty$ , each strength of schedule will tend to 0 expected wins, whereas as in the limit as  $\beta_0 \rightarrow \infty$  the expected number of wins will tend to the number of games played. As we are interested in distinguishing between teams on the borderline for receiving an at-large bid, it is most

appropriate for  $\beta_0$  to be chosen to be similar in value to the value of  $\beta$  for such a team. In practice our approach is to choose  $\beta_0$  such that exactly 34 non-conference champions have more wins than the expected number of wins given by their strength of schedule. These 34 teams are thus the ones which this approach suggests should be given at-large bids. In practice, small variations around this value make little impact on the rankings of the teams.

## 4 Results

We now analyse data from both the 2007/08 and 2008/09 regular seasons. In both cases we obtained results of all matches between first division teams from ESPN's website (<http://espn.go.com/mens-college-basketball/>). First we present results of the analysis for the 2008/09 season, with a focus on the issues we have raised above about how informative are different summaries of a team's results. We then give the results and interpretation for the two seasons, focussing specifically on the choice of teams for at-large bids. Full results from these two seasons are available from

<http://www.maths.lancs.ac.uk/~fearnhea/Basketball.html>

### 4.1 Summaries of Teams' Results

For the 2008/09 season the posterior mean estimates of the  $\beta$ s ranged from  $-22.7$ , for the N.J.I.T. to  $23.1$  for the University of North Carolina. The size of home advantage,  $\gamma$ , was estimated at 3.5 points, and the standard deviation of the residuals was 10 points. The latter two are similar to the estimates from the 1999-2000 season in Harville (2003).

Rank	Team	$\beta_j$	Win Prob. (1)			Win Prob (2)
			H	N	A	N
1	North Carolina	23.1	0.22	0.13	0.07	0.00
19	Georgetown	13.6	0.57	0.43	0.30	0.01
24	Kansas St.	12.7	0.60	0.46	0.33	0.01
37	LSU	11.2	0.66	0.52	0.39	0.01
68	Dayton	7.0	0.80	0.68	0.55	0.03
109	Rutgers	2.2	0.90	0.83	0.73	0.08
314	Alcorn State	-20.9	1.00	1.00	1.00	0.82

Table 1: Subset of results for the 2008/09 season. Rank gives the rank, in terms of the posterior mean of the  $\beta_j$ s, *among teams that did not win their conference championship*. Win Prob. gives the predicted win probability of a fictitious team: (1) corresponds to a team with  $\beta_0 = 11.8$ , and (2) to one with  $\beta_0 = -11.8$ . Columns H, A, and N correspond to a home, away and neutral match for the fictitious team.

A small selection of teams, and corresponding estimated  $\beta_j$ s are given in Table 1. This table also gives the predicted probability that two fictitious teams would win a match against each of these teams. The first of these teams is chosen to be of a standard comparable to a team on the borderline of getting a bid – and is the standard of team we use in calculating strength of schedules below. The second team is chosen to be of a standard comparable to one of the weaker division I teams.

The results in Table 1 help us to quantify how the strength of schedule depends on the ability of the fictitious team. For the pair of two match schedules suggested in Section 2, a strong fictitious team ( $\beta_0 = 11.8$ ) would expect to win

1.22 games against North Carolina and Alcorn State, and 1.66 games if they played Rutgers twice. By comparison a weaker team ( $\beta_0 = -11.8$ ) would win on average 0.82 matches for the first schedule, but only 0.16 for the second. Thus the ability of the team we consider can even effect not only the absolute difficulty of each schedule, but even which schedule is easier.

The results in Table 1 also highlight how the importance of home advantage depends on the relative ability of the two teams. For comparable team, such as the stronger fictitious team and Kansas State, playing at home as opposed to playing away can increase the probability of winning by over a quarter. For very disparate teams, such as the stronger fictitious team and Alcorn State, it has negligible effect. By comparison, one disadvantage of the way that the RPI accounts for home-advantage is that it quantifies its effect equally for all matches.

From Table 1 we see the problem of using the estimates of the  $\beta_j$ s to rank teams for inclusion in the national tournament. Neither Georgetown nor Kansas State recieved bids for the tournament in 2009, whilst Dayton did. In total these rankings gave 9 teams in the top 34 non-conference champions that did not receive bids to the tournament. The reason for this disparity is that the estimates of the  $\beta_j$ s are based on point-differences in matches and not just the results of those matches. So teams like Georgetown had few wins given their net point-difference in the matches they played.

Finally, we also see that there is a much larger difference between the estimate of  $\beta_j$  for North Carolina and Georgetown, as compared to that between Georgetown and either LSU or even Dayton. This highlights the problem of summaries based on win-loss records against top-50 teams, as it ignores the substantial difference in difficulty of a match against a team like North Carolina,

as opposed to one like Georgetown. The biggest differences in ability of teams is within roughly the top-10 teams, and there is comparably little difference between teams ranked just within or just outside the top-50.

## 4.2 2008/09 Season

We now focus on the estimated strength of schedules for teams from the 2008/09 season, and the information they give about who deserved at-large bids to the national tournament. We calculated strength of schedule for all teams, and we give results for a subset of teams in Table 2. Here we focus on teams ranked by the difference in their actual win percentage and the win percentage given by their strength of schedule. We excluded conference champions, and give results for the teams close to position 34 – the borderline for receiving an at-large bid.

As described in Section 3, we fixed the ability of the fictitious team to such that 34 non-conference champions won more matches than their strength of schedule. This corresponded to  $\beta_0 = 11.8$ . We also quantified the uncertainty in the estimates of the strength of schedule, through estimating the posterior variance of the predicted win probability for each match. To summarise these, we calculated the average variance in the estimate of the strength of schedules for all teams. This average variance is 0.088. This suggests that estimates of strength of schedules are accurate to within plus or minus 0.6, based on twice the corresponding standard error.

The 34 non conference champions who won more games than their strength of schedule (see Table 2) contain two teams not given at-large bids to the national championship. These were St. Mary's and San Diego State, with Michigan and Wisconsin receiving bids instead. The choice between San Diego State and

Rank	Team	W-L	SOS	Rank	Team	W-L	SOS
27	Minnesota*	21-10	20.2	35	Michigan*	19-13	19.1
28	Saint Mary's	24-6	23.3	36	Tulsa	23-10	23.3
29	Boston College*	22-11	21.5	37	Creighton	26-7	26.5
30	Tennessee*	21-12	20.5	38	Penn State	22-11	22.6
31	Dayton*	26-7	25.8	39	Florida	23-10	23.6
32	Maryland*	20-13	19.8	40	Miami (FL)	17-12	17.6
33	San Diego St.	22-9	21.8	41	Baylor	19-14	19.7
34	Arizona*	19-13	18.9	42	Wisconsin*	19-12	19.7

Table 2: Results for 2008/09 season. Rank is based on the difference of actual win percentage and win percentage of the strength of schedule (SOS). In calculating the rank we have exclude all conference champions. Thus teams our results on the left should have received an at-large bid, and those on the right should not have. Teams who actually received an at-large bids are marked by \*. SOS is given in terms of the predicted number of wins.

Michigan was very close, the former's strength of schedule was 0.2 less than the number of games won, and the latter's was 0.1 more. These are both comfortably within the margin of error of the estimates of the strength of schedule, which was 0.6.

However there appears to be clear evidence that St. Mary's should have been preferred to Wisconsin for an at-large bid. The omission of St. Mary's from the national championship could be explained by an injury to one of their players, Paddy Mills, near the end of season, and resulting uncertainty over whether he would be fully fit for the championship. However, even allowing for this, the results in Table 2 suggest that teams other than Wisconsin should have been given a bid instead.

### **4.3 2007/08 Season**

We now repeat the analysis for the 2007/08 season and results are given in Table 3. We again estimate the margin of error of the strength of schedules as twice the estimated standard error, and this is 0.6.

This time there are 4 teams who have won more games than their SOS, but who were not given at-large bids. All four teams had the difference between their number of wins and their strength of schedule at or within our estimated margin of error. Of the four teams who received at-large bids instead, Mississippi State were on the borderline: ranked 35th among non-conference champions and with their number of wins equal to their strength of schedule. Kentucky had performed poorly early in the season, due to injury problems, but had performed much better during the second half of the season (12-4 in the SEC). It is likely that the early season injury problems were taken into account in the decision to

give them an at-large bid.

However our results suggest that the other two teams, South Alabama and St. Joseph's were fortunate to receive bids. In particular St. Joseph's won 21 matches with a strength of schedule of 22.7, suggesting that they required 2 extra wins to warrant an at large bid. To look into this further we compare the schedules of St. Joseph's with that of Arizona State, who won 19 matches with a strength of schedule of 18.4 but were not given an at-large bid. The schedules, together with the win probabilities that are used to calculate the strength of schedule are given in Table 4.

The argument against Arizona State is that they had "RPI in the 80s, nonconference SOS in the 300s, a 9-9 league record and 5-10 in its last 15 matches" Glockner (2008). By comparison St. Joseph's RPI rank was 44 (see <http://kenpom.com/rpi.php>). Arizona State's record against top-50 RPI teams was 5-7, and St. Joseph's was 5-6.

One of the main aspects of the argument for preferring St. Joseph's is their better RPI rating. However, Arizona State's RPI is severely affected by the matches they played against weak teams. If you remove the 10 easiest matches from their schedule, then their RPI would improve to about 50th – despite the fact they won all 10 of these matches. Also whilst St. Joseph's record against top-50 RPI teams is marginally better, looking at Table 4 we can see that on average their matches against top-50 opponents were easier than Arizona State's. The overall expected number of wins for St. Joseph's against their top-50 opponents is 5.3, whereas Arizona State's is just 3.7.

Rank	Team	W-L	SOS	Rank	Team	W-L	SOS
27	Davidson*	25-6	24.2	35	Mississippi St.*	22-10	22.0
28	Oregon*	18-13	17.2	36	New Mexico	24-8	24.1
29	Arizona St.	19-12	18.4	37	Mississippi	21-10	21.3
30	Arkansas*	22-11	21.6	38	Syracuse	19-13	19.6
31	Nebraska	19-12	18.7	39	South Alabama*	25-6	25.8
32	Villanova*	20-12	19.7	40	Creighton	21-10	21.8
33	Illinois St.	23-9	22.8	45	Kentucky*	18-12	19.2
34	Ohio State	19-13	18.9	50	St. Joseph's*	21-12	22.7

Table 3: Results for 2007/08 season. Details are the same as in Table 2.

## 5 Discussion

We have presented a method for calculating a team's strength of schedule in terms of an expected number of wins. This gives a benchmark performance with which to compare a team's actual results, and has been constructed so that if a team wins more games than their strength of schedule, that is evidence that they deserve a bid to the national tournament.

We think this approach has advantages over existing methods for ranking teams, or summarising their performances. The strength of schedule is interpretable, and enables a quick comparison between the performance of different teams after taking account of the schedules that they have each played. It enables teams to be ranked in terms of their win-loss records, as opposed to methods that are influenced by winning margins. Furthermore, we can assess uncertainty in our estimate of the strength of schedule, so we can not only highlight the 34 teams most deserving of bid but also suggest which teams have demonstrated clear

Arizona State			St. Joseph's		
Opponent	Res.	Prob.	Opponent	Res.	Proc.
at UCLA*	L	0.08	at Xavier*	L	0.18
at Stanford*	L	0.17	vs Xavier*	W	0.30
at Washington St.*	L	0.18	at Duquesne	L	0.41
at Arizona*	W	0.23	at Creighton*	L	0.42
at USC*	L	0.25	at Massachusetts*	W	0.43
UCLA*	L	0.27	at Rhode Island	W	0.44
at Oregon	L	0.29	Xavier*	W	0.44
at California	W	0.38	at Temple*	W	0.45
vs USC*	L	0.38	at Dayton	L	0.47
at Washington	W	0.39	vs Syracuse*	L	0.48
Stanford*	W	0.43	at Charlotte	L	0.51
Xavier*	W	0.44	Gonzaga*	L	0.54
Washington St.*	L	0.45	at Siena	W	0.60
vs Nebraska	L	0.49	vs Temple*	L	0.60
Arizona*	W	0.51	Villanova	W	0.69
vs Illinois	L	0.53	Massachusetts*	W	0.72
USC*	W	0.53	Temple*	L	0.74
Oregon	W	0.59	vs Penn State	W	0.75
California	L	0.67	at George Washington	W	0.76
Washington	L	0.69	14 other matches	W11-L3	12.7
vs LSU	W	0.72			
10 other matches	W10-L0	9.7			
Total	W19-L12	18.4	Total	W21-L12	22.7

Table 4: 2007/08 schedules and results for Arizona State and St. Joseph's, together with the expected win probability that is used in calculating each team's strength of schedule. Away matches are denote by 'at', and neutral matches by 'vs'. Opponents within the RPI top-50 are shown denoted by \*. For ease of presentation we have ordered matches by their difficulty, and grouped together all matches where the win probabiity was greater than 0.8.

evidence that they warrant a bid, and which teams are on the borderline of selection. For teams whose season has been affected by injuries, such as Kentucky in 2007/08, we can break the strength of schedule down into subsets of matches and thus look directly at how a team has performed at different stages of the season. As such we believe this strength of schedule would help inform decisions as to which teams are given bids to the national tournament.

## References

Carlin, B. P. (1996). Improved NCAA basketball tournament modeling via point spread and team strength. *The American Statistician*, 50:39–43.

cheap ticket to the dance: systematic bias in college basketball's rating percentage index, A. (2007). S sanders. *Economics Bulletin*, 4:1–7.

Glockner, A. (2008). Final day for the bubble teams to make case. Internet Article.

Harville, D. A. (2003). The selection or seeding of college basketball or football teams for postseason competition. *Journal of the American Statistical Association*, 98:17–27.

Harville, D. A. and Smith, M. H. (1994). The home-court advantage: how large is it and does it vary from team to team? *The American Statistician*, 48:22–28.

Stefani, R. T. (1977). Football and basketball predictions using least squares. *IEEE transaction on systems, man, and cybernetics*, SMC-7:117–121.

Stefani, R. T. (1980). Improved least squares football, basketball and soccer predictions. *IEEE transaction on systems, man, and cybernetics*, SMC-10:116–123.