7-2009

NCAA Tournament Games: The Real Nitty-Gritty

B. Jay Coleman
University of North Florida, jcoleman@unf.edu

Allen K. Lynch

NCAA Tournament Games: The Real Nitty-Gritty

Jay Coleman, *University of North Florida*
Allen K. Lynch, *Mercer University*

NCAA Tournament Games: The Real Nitty-Gritty

Jay Coleman and Allen K. Lynch

Abstract

The NCAA Division I Men's Basketball Committee annually selects its national championship tournament's at-large invitees, and assigns seeds to all participants. As part of its deliberations, the Committee is provided a so-called "nitty-gritty report" for each team, containing numerous team performance statistics. Many elements of this report receive a great deal of attention by the media and fans as the tournament nears, including a team's Ratings Percentage Index (or RPI), overall record, conference record, non-conference record, strength of schedule, record in its last 10 games, etc. However, few previous studies have evaluated the degree to which these factors are related to whether a team actually wins games once the tournament begins. Using nitty-gritty information for the participants in the 638 tournament games during the 10 seasons from 1999 through 2008, we use stepwise binary logit regression to build a model that includes only eight of the 32 nitty-gritty factors we examined. We find that in some cases factors that receive a great deal of attention are not related to game results, at least in the presence of the more highly related set of factors included in the model.

KEYWORDS: binary logit, stepwise, committee decision, performance metrics

1. INTRODUCTION AND BACKGROUND

During March and early April of each year the National Collegiate Athletic Association (NCAA) holds its Division I Men's Basketball National Championship Tournament (commonly referred to as simply the NCAA Tournament). Approximately 31 of the Tournament's 65 participants receive automatic bids by winning their respective conference championships, which are nearly exclusively identified through post-season conference tournaments. The remaining 34 teams gain entry by receiving at-large invitations from the NCAA Division I Men's Basketball Committee (herein referred to as the Committee). The Committee is also charged with assigning a seed (from 1 to 16) to each team and placing each team in one of four regional brackets. Two of the 65 teams participate in a "play-in game" to earn one of the 16 seeds.

Although the Committee's work takes place out of the public eye, the general process by which the Committee makes its decisions is published by the NCAA. Part of the selection and seeding process is the availability of a so-called "nitty-gritty report" for any team under consideration (NCAA 2005a, 2005b). This report contains a large amount of data capturing various aspects of each team's performance over its season.   For example, each team's overall, conference, non-conference, and road records are listed, as well as data on the performance of its opponents designed to capture the strength of each team's schedule. Perhaps the most important variable in the report, and the means by which teams and their conferences are rated, ranked, and categorized, is the Ratings Percentage Index (or RPI), a metric devised by the NCAA to aid in the assessment of team strength (discussed below).

Although the NCAA's published "principles and procedures for establishing the bracket" for the 2007 and 2008 Tournaments (NCAA 2006a, 2007) did not specifically mention the nitty-gritty report, the documents still contained a discussion of the RPI, as well as a note that various resources are made available to the Committee, including items such as Division I record, home and away records, non-conference records, etc., which appear rather consistent with the list of items previously published as being elements of the nitty-gritty report. Comments by recent Committee chairs have also made mention of similar items (Littlepage 2006a; O'Connor 2008b). Additionally, nitty-gritty reports have been mentioned by name in articles discussing recent mock selection meetings the NCAA holds to illustrate the process to media members (Henderson 2008). Moreover, in 2006 the NCAA themselves for the first time started releasing the RPI rankings to the public (NCAA 2006b), underscoring the continued presence of this measure as a decision aid. Additionally, sports media and noted college basketball web sites (such as CollegeRPI.com and kenpom.com) have continued to publish and/or remark on the RPI as well as many other traditional nitty-gritty

report items, emphasizing their substantial and continuing visibility to those who follow the game.

1.1. The Ratings Percentage Index

A large portion of the nitty-gritty report information is in some way based on a version of the RPI. The NCAA computes different forms of the RPI for each team, including its overall RPI (which is generally referred to as simply "the RPI," and which is the version that is widely reported and discussed by media and fans as the Tournament approaches each year), and its non-conference RPI. (A third form, the conference RPI, is also computed and reported by some sources such as CollegeRPI.com.) Each of these is the weighted average of three parts. The first part (carrying a 25% weight) differs across each version. In the overall RPI, the first part is the team's winning percentage against all of its Division I opponents. The first part of the non-conference RPI is the team's winning percentage against only its non-conference opponents; and the first part of the conference RPI is the team's winning percentage against its conference opponents. The second part of all three RPI versions is the overall winning percentage of those opponents included in the computation of the first part, and it receives a weight of 50%. The third part of all three RPI versions is the overall winning percentage of those same opponents' opponents (weighted by 25%).[1]

Thus, the conference RPI and non-conference RPI reflect a team's record in and out of conference, respectively, and the strength of those teams played in and out of conference, respectively. The NCAA also computes strength of schedule metrics for each team, computed by weighting the second component of the respective RPI by two-thirds, and the third component in the respective RPI by one-third (i.e., the same relative weights used in the RPI itself).

In 2005, the Committee changed the inputs into just the first part of the various RPI versions (note that the strength of schedule components of the RPI calculation were unaffected by the change). A win on the road was counted as 1.4 wins, a win at home was counted as 0.6 of a win, a loss at home was counted as 1.4 losses, and a loss on the road was counted as 0.6 of a loss. This adjustment caused great debate among many in the media and throughout college basketball, as the resulting RPI rankings were changed significantly (Cohen 2006).

_____

[1] As noted in West (2006, 2008), the NCAA's selection of these particular weights for the various parts of the RPI appears arbitrary. However, when the RPI was first conceived by the NCAA in the late 1970's (leading ultimately to its first use in 1981), statisticians from Stanford were reportedly involved, as was the NCAA's director of statistics at the time, and the computation went through multiple early versions. The weighting scheme for the three elements formerly was 20-40-40, respectively, as opposed to the current 25-50-25, until the NCAA decided that the heavier weight on the third component sometimes created "false impressions, depending on the strength of a particular team's conference" (Brown, 1999).

However, in each year since the change (through 2009), the "old" overall RPI ranking was as good or better than the "new" overall RPI ranking in predicting who received the Committee's at-large bids.[2] As a result, it is the "old" version of the RPI, in which all wins and losses are counted as one win or one loss regardless of location, which is the basis for the study discussed here – and it was the version that was in use by the NCAA to generate the nitty-gritty reports during most of the time frame covered by our study.[3]

1.2. Previous Studies

Given the secrecy of the Committee's discussions during its at-large team selection process, Coleman and Lynch (2001) attempted to model the Committee's decisions by comparing its selections from 1994 through 1999 to nitty-gritty report information for each team in those seasons. They identified a six-factor model that was able to accurately predict at-large bids with 94% accuracy in-sample, and which has been nearly as successful in predicting bids out-of-sample for 2000 through 2008 (Coleman 2008).

Sanders (2007) examined the construction of the RPI and whether it contained systematic conference bias. Jing and Cox (2008) used 2001-2004 nitty-gritty report information similar to that used by Coleman and Lynch, plus the rankings according to *USA Today's* Jeff Sagarin, to examine the efficacy of machine learning methods at predicting the Committee's decisions.

However, none of the above work examined the relationship between nitty-gritty factors and game results. In contrast, the relationship between the Committee's assigned seeds and game results has been a focus of earlier research. Schwertman, McCready, and Howard (1991), Schwertman, Schenk and Holbrook (1996), Stern and Mock (1998), Smith and Schwertman (1999), Boulier and Stekler (1999), and Caudill (2003) examined the seeds of the teams involved in each game, or some derivation thereof, as the predictor(s) in determining the probability of whether a particular seed wins a given Tournament game and/or its respective regional championship. With the exception of Caudill (2003), each of the probability models developed by these authors simply predicts that the higher seed will win each game (i.e., the predicted winner would be the same for each game, but the probability of winning would change depending on the model used). However, Caudill's approach would have modified such a prediction in only four of the 840 games he examined, by incorporating the difference in seeds

_____

[2] From 2005 through 2009, the "new" RPI missed 3, 5, 6, 4, and 3 at-large bid predictions, respectively, whereas the "old" RPI missed 3, 2, 6, 4, and 2 at-large bid predictions in those years (new RPI rankings were taken from Palm, 2009).

[3] Stated otherwise, all the analyses reported in this study employed the "old" RPI for every team, regardless of year.

(like the earlier authors) as well as the seed of each team. Carlin (1996) took a different approach to using seed information. He used seeds and Jeff Sagarin's pre-Tournament team strength ratings to predict what the pre-Tournament betting point spreads likely would have been for all possible second, third, and fourth round match-ups of the 1994 Tournament teams. Using the actual first round betting point spreads, along with his later-round predicted spreads, and an assumed standard deviation of 10 points for both, he estimated the probability of each team winning its regional championship.

Harville (2003), Stern et al. (2004), Kvam and Sokol (2006), and West (2008) are the only published works to make mention of the predictive accuracy of the overall RPI, the only nitty-gritty item previously so examined. Harville (2003) compared the accuracy of the RPI, the seeds, and the betting line for the games of the 2000 Tournament, as part of an evaluation of his ranking model. Stern et al. (2004) mentioned the performance of the RPI as part of an overall discussion of ranking systems. Kvam and Sokol (2006) presented a logistic regression / Markov chain (LRMC) model designed to predict the results of Tournament games, and compared its performance over 2000-2005 against various benchmarks, including the RPI and the seeds. West (2006, 2008) presented an ordinal logistic regression (OLRE) model for predicting the number of tournament games a team will win, using as predictors a team's overall winning percentage, Sagarin strength of schedule, number of wins against the Sagarin top 30, and total point differential in pre-Tournament games. West discusses the RPI and its various forms in both articles, and in the latter compares his ordinal logistic regression model to the RPI in terms of the capability to predict upsets.

2. RESEARCH QUESTION AND CONTRIBUTION

Despite the nitty-gritty report's apparently significant role in the process, and despite the media and fan attention that much of its content receives, few previously published studies have determined the relationship between the various contents of the nitty-gritty report and the results of Tournament games. Therefore, it is this research question that we attempted to address.

West's (2006, 2008) use of the overall winning percentage in his OLRE model, and his above-mentioned comparison to the RPI, represents the closest such study in the literature. We extend West through the consideration of numerous additional nitty-gritty factors (32 in all), and the use of more years (10) in building our models (West (2006, 2008) considered four years of data when model-building). Our study therefore addresses two of the limitations that West stated regarding his analysis (West 2006, 2008). We also employ a tournament *game* as the unit of observation, whereas West considered the *team* as the unit of

observation (and used the number of wins by each team as the dependent variable). Doing so allows our analysis to control for the strength of the opponents actually played by each team. Stated otherwise, we in effect control for the seeding decisions of the selection committee, a source of variation that could impact results using West's OLRE methodology.

If the committee happens to weight some regions with higher quality teams than other regions, and/or if the committee assigns inordinately low (or high) seeds within a region to teams that deserve better (or worse) seeding, some teams may end up with fewer (or more) actual wins in the tournament than their strengths would warrant. To the extent that the committee has distributed (or will distribute) teams inequitably across and/or within regions, the number of wins by each team is affected. Use of a game instead of the team as the unit of measure mitigates this concern.

In addition, our research objective as stated at the outset of this section is by extension similar to that of West (2006, 2008): to identify information that could potentially be of use by the Committee in its seeding and regional balancing decisions for those teams that make the Tournament, whether it be by automatic or at-large bid. We seek to identify a subset of the factors on the NCAA's own report that is highly related to Tournament wins. Thus, our research objective and our findings are relevant to the Committee's decision-making process to the extent that the Committee is interested in factors and models that reflect the future performance of teams once they are in the Tournament. However, like in West, our objective does not extend to the Committee's at-large selection process. As stated by West regarding his models, since the research reported below is "based on the patterns of success for selected teams in the *previous* years [meaning that they are] fitted using historical data for teams selected for the tournament, [our results do] not apply to teams that are not selected for the tournament" (West 2006).

3. DATA AND VARIABLES

In order to compute the values of the nitty-gritty factors that were used as *predictors* in our model, for the 10 seasons of 1999 through 2008 we collected all regular season and conference tournament game results – i.e., all games that *preceded* the NCAA Tournament in each year – from Ken Pomeroy, a contributor to ESPN.com who maintains a popular website (kenpom.com) on college basketball. Using only the games involving two Division I teams, we computed or collected 32 different statistics for each team for the respective season (see Table 1). Most of these factors are estimates of nitty-gritty report statistics or derivations of such estimates, and many are comparable to the factors analyzed by Coleman and Lynch (2001) and/or included among factors highlighted by

Table 1. Nitty-gritty report factors considered for each team.

Overall RPI
Overall strength of schedule
Conference RPI
Conference strength of schedule
Number of conference wins
Number of conference wins above a 0.500 record (0 if record 0.500 or below)
Number of conference losses below a 0.500 record (0 if record 0.500 or above)
Non-conference RPI
Non-conference strength of schedule
Number of non-conference wins
Number of non-conference games above a 0.500 record (0 if record 0.500 or below)
Number of non-conference games below a 0.500 record (0 if record 0.500 or above)
Number of road wins
Number of road wins above a 0.500 record (0 if record 0.500 or below)
Number of road losses below a 0.500 record (0 if record 0.500 or above)
Number of neutral court wins
Number of neutral court wins above a 0.500 record (0 if record 0.500 or below)
Number of neutral court losses below a 0.500 record (0 if record 0.500 or above)
Wins vs. teams ranked 1-25 in RPI
Number of wins above a 0.500 record against teams ranked 1-25 (0 if record 0.500 or
    below)
Number of losses below a 0.500 record against teams ranked 1-25 (0 if record 0.500 or
    above)
Wins vs. teams ranked 26-50 in RPI
Number of wins above a 0.500 record against teams ranked 26-50 (0 if record 0.500 or
    below)
Number of losses below a 0.500 record against teams ranked 26-50 (0 if record 0.500 or
    above)
Wins vs. teams ranked 51-100 in RPI
Number of wins above a 0.500 record against teams ranked 51-100 (0 if record 0.500 or
    below)
Number of losses below a 0.500 record against teams ranked 51-100 (0 if record 0.500 or
    above)
Number of losses against teams ranked lower than 100 in RPI
Number of wins in the last 10 games
Rank of team's conference in that season, according to the mean non-conference RPI of
    all teams in the conference
Binary variable reflecting whether a team won its regular season conference
    championship (or co-championship)
Binary variable reflecting whether a team received an automatic bid to the Tournament
    (almost exclusively achieved through winning its conference tournament)

Pomeroy and Jerry Palm of CollegeRPI.com, among others (e.g., Zuchowski (2005), Henderson (2008)). Included are relatively well-known and oft-discussed nitty-gritty items such as each team's overall RPI, its conference record, its record against teams ranked at various positions in the RPI, its record in its last 10 games (although this was changed by the NCAA in 2006 to record in the last 12 games, we used the last 10 games throughout our analysis), its conference's strength ranking among all conferences competing that season, and its non-conference strength of schedule. The non-conference strength of schedule has been specifically highlighted by recent Committee chairs as being particularly important (Bowlsby 2005; Littlepage 2006a; O'Connor 2008b; Prisbell 2008). We avoided some nitty-gritty metrics that were linear combinations of others, such as number of overall wins (the sum of conference and non-conference wins) and overall winning percentage (a function of overall RPI and overall strength of schedule).

In several cases we used raw wins and losses information to derive alternative variables to try to capture the way in which Committee members were apt to view the data. For example, instead of using (as Coleman and Lynch (2001) did) a single win-loss record variable such as wins minus losses (or winning percentage) against various groups of opponents (e.g., overall, conference, non-conference, etc.), we derived separate variables reflecting the number of games a team finished above or below a 0.500 record against that same group. This was done under the presumption that the Committee likely weights winning records quite differently than losing records – and likely thinks along the traditional sports perspective of games above or below 0.500 as opposed to percentage points above or below 0.500 – and thus we allowed for a possible break point in the relationship.

We also added binary dummy variables reflecting whether the team won its conference tournament or its regular season conference championship, each of which – whether actually included in the report – are known and considered relevant by Committee members (Bowlsby 2005; Littlepage 2006b). Finally, for comparison's sake, we also collected the Tournament seed that was assigned to each team by the Committee as well as the final pre-Tournament ranking of each team according to Jeff Sagarin.

Additionally, we obtained the participating teams and the final scores for all 638 NCAA Tournament games from 1999 through 2008 (including the play-in games) from Pomeroy. These 638 games represented our unit of analysis when model-building. Using a similar approach to that used by Carlin (1996), Schwertman et al. (1996), Smith and Schwertman (1999), Boulier and Stekler (1999) and Caudill (2003) when analyzing the predictive information contained within the seeds, to determine the value of each *predictor* for each observation we computed the difference in the respective values of each of the 32 pre-

Tournament statistics for the two teams in each game. All differences for each game were computed from the perspective of the higher ranked team in RPI (i.e., the value for the higher ranked team was always taken first in the calculation). For example, if a team with an RPI ranking of 32 had won 10 conference games, and played a team with an RPI ranking of 48 that had won 8 conference games, then the difference in conference games won would have been computed as 10 – 8 = 2. Although the authors cited above computed differences for each game from the perspective of the higher *seed* (i.e., the value for the higher seed was always taken first in the calculation), our approach allowed the data set to include the play-in games and games involving identical seeds, as occasionally happens during the Final Four (the last two rounds of the Tournament). (The earlier authors analyzed only the games in the first four rounds – i.e., the games within each of the four regions – and not the three games of the Final Four.) However, our data did include one game in which the RPIs of the two teams were identical – we chose the winner of that game as the first team in the difference calculations.

This process generated 32 corresponding potential predictor variable values for each of the 638 NCAA Tournament games in 1999-2008. Similar to Schwertman et al. (1996), Smith and Schwertman (1999), Boulier and Stekler (1999, 2003) and Caudill (2003), the dependent variable was constructed as a binary variable, in our case representing whether the higher ranked team in RPI won the Tournament game in question (1 if it won, 0 otherwise). Since the higher ranked team in RPI won 444 of the 638 Tournament games studied (including the one game between teams tied in RPI), 444 observations contained a value of 1, and the remaining 194 contained a value of zero.

As control variables we constructed five binary dummy variables representing the round in which the respective Tournament game was played (the first round was the omitted category). The Tournament takes place over the course of six rounds in which the field is progressively narrowed from 64 teams (not counting the play-in game) in the first round to a single champion at the conclusion of the sixth round. These variables were included to examine whether the round in which the game was played had an effect on the probability that the team with the higher RPI would win the game. We also interacted these five dummies with variables otherwise found to be significant predictors, to determine if the relationship of these factors to the likelihood of winning a Tournament game changed based on the round in which that game was played.

Finally, to allow for more straightforward assessment of the relative importance of various factors in predicting the probability of winning a Tournament game, prior to model fitting we standardized all non-binary predictor values by converting each to a z-score.

4. MODEL ESTIMATION

Given the binary nature of the dependent variable, a binary logit analysis was selected as an appropriate option for estimating the model. The predicted value generated by the logit model was converted into a probability of the higher ranked team winning the game. If the estimated probability (obtained from the model) of the higher ranked team winning was greater than 0.50 and the higher ranked team in reality won, then the model was deemed to have correctly predicted that game. Similarly, the model was deemed to have correctly predicted the game if the estimated probability obtained from the model was less than 0.50, and the higher ranked team lost the game. Although obviously an argument could be made that a predicted value of exactly zero (or 50% probability, or in essence a predicted tie) would by definition be a model error, as ties are not a possibility in college basketball, the chance of a fitted value of zero is very small, and did not result for the model discussed below.

Clearly, there are numerous variables in Table 1 that are correlated – e.g., the strength of schedule metrics were direct components of their corresponding RPI's – and thus using all predictors in the same model would likely lead to misleading interpretations, and represent an over-specification of the model. Thus, to specify the model we used a stepwise logit procedure employing PROC LOGISTIC in SAS Version 9.1 (SAS 2004), using a 0.20 (two-tailed) significance level for both variable entry and exit. Consistent with the recommendations of Hosmer and Lemeshow (1989, p. 108), we chose this rather unrestrictive significance level to encourage inclusion of important variables. The result of this stepwise procedure was named Model 1.

In the stepwise process that generated Model 1, we omitted the round dummies and any interactions of the round dummies with our nitty-gritty predictors.[4] However, in a second round of stepwise modeling we included all significant predictors from Model 1 as potential predictors to be selected, plus the five round dummies and the interactions between these five dummies and every nitty-gritty factor in Model 1.

Also, given the attention granted the RPI (in its various forms) as the most prominent nitty-gritty factor, in a third round of stepwise modeling we forced each of the three RPI's – overall RPI, conference RPI, and non-conference RPI – individually and (in the case of the conference and non-conference RPI) collectively into the model, and allowed stepwise to choose the remaining predictors to add to the model.

---

[4] One rationale for this was that with five round dummies and 32 potential predictors, inclusion of all interaction terms (plus the five round dummies themselves) would have necessitated the addition of 165 potential predictors to the stepwise procedure.

Finally, we also assessed the statistical significance associated with adding the difference in seeds to Model 1. This was done to determine whether our model simply captured the same information the Committee captured in its seeds. If the difference in seed was significant in the presence of the nitty-gritty factors in the model, it would imply that the deliberative process of the committee generated more predictive information than was captured in the nitty-gritty factors alone.

5. ASSESSING MODEL FIT AND PREDICTIVE POWER

Each model's fit was assessed in several ways. We performed Hosmer-Lemeshow goodness-of-fit tests, and (like West (2006, 2008)) examined the Nagelkerke adjusted (max-rescaled) R-square value. We also examined the Akaike Information Criterion and Schwarz Criterion for each model. To put fit in terms of game results, we computed the percentage of the 638 in-sample games that were correctly classified by each model. To provide some context for this latter value, we compared the model's results over these 638 games to those of the higher seeds, to those yielded by the Sagarin rankings, and – given the RPI's prominence to the college basketball public – to the teams with the higher RPI.

Finally, to examine each model's predictive power, we assessed how well it predicted the winners of the 63 games played during the 2009 NCAA Tournament. Moreover, and similar to West (2006, 2008), we also computed the predicted number of tournament wins by each team in 2009 according to each model, and compared those predicted values to the actual number of wins by each team using a sum of squared errors (SSE). We then compared this SSE to the SSE generated for the 2009 Tournament for each of the three models presented in West (2006, 2008): an unadjusted OLRE, an adjusted OLRE, and a Bradley-Terry simulation (the 2009 results from these three models were provided in West (2009)). We also compared the 2009 predictions to those from Sagarin, from the RPI alone, and from the seeds, as well as to the performance of the Las Vegas betting favorites (Covers.com, 2009) and Kvam and Sokol's LRMC method noted earlier (Kvam and Sokol, 2009; Sokol and Nemhauser, 2009a, 2009b). Along with West (2006, 2008), the LRMC model represents one of the few Tournament prediction methods published in the academic literature, for which predictions and/or rankings are posted online. It also has been shown to be a better performer than other leading methods (including Sagarin's) during 2000 – 2006.

To determine each model's predicted number of games won by each team in 2009, we first used the predicted values from the model to determine the 64 x 63 = 4032 separate probabilities that each of the 64 teams would win a game played against every one of the other 63 teams in the Tournament. The predicted

probability that team i would defeat team j, if team i was *higher* ranked in RPI, was computed as

$$\hat{\pi}_{ij} = \frac{\exp(\alpha + (x'_i - x'_j)\beta)}{1 + \exp(\alpha + (x'_i - x'_j)\beta)} \tag{1}$$

where x_i is the vector of pre-Tournament metrics included in the model for team i, x_j is the vector of corresponding pre-Tournament metrics included in the model for team j, and β is the vector of coefficients for those corresponding predictors in the model. The predicted probability that team i would defeat team j, if team i was *lower* ranked in RPI, was computed as

$$\hat{\pi}_{ij} = 1 - \frac{\exp(\alpha + (x'_j - x'_i)\beta)}{1 + \exp(\alpha + (x'_j - x'_i)\beta)} \tag{2}$$

We then constructed a 64 x 6 cumulative probability table, with each entry representing the probability that the team in that row would proceed *through* the round in that column. These cumulative probabilities accounted for not only the chance that the team in question would win a game in that round, but also the likelihood that the team in question and the possible opponents in that round even made it that point in the Tournament.[5] The cumulative probability that team i would win in round 1, where j was i's round 1 opponent, was simply the predicted probability of i defeating j:

$$\hat{\pi}_{i1}^c = \hat{\pi}_{ij} \tag{3}$$

The cumulative probability that team i would win in round 2, where k is an element of the set of two possible teams that team i could play in round 2, reflected the probability that i would make it through the first round, the probability that it would defeat each of the two teams it could possibly play in the second round, and the probabilities that each of those teams would make it through the first round:

$$\hat{\pi}_{i2}^c = \hat{\pi}_{i1}^c \times \sum_{\forall k} (\hat{\pi}_{ik} \times \hat{\pi}_{k1}^c) \tag{4}$$

Similarly, the cumulative probability that team i would win in round 3, where m is an element of the set of four possible teams that team i could play in the round 3, was

$$\hat{\pi}_{i3}^c = \hat{\pi}_{i2}^c \times \sum_{\forall m} (\hat{\pi}_{im} \times \hat{\pi}_{m2}^c) \tag{5}$$

---

[5] Entries in each row (i.e., for each team) were monotonically decreasing from left to right (i.e., from round 1 to round 6).

The cumulative probabilities $\hat{\pi}_{i4}^{c}$, $\hat{\pi}_{i5}^{c}$, and $\hat{\pi}_{i6}^{c}$ were similarly and sequentially computed for each team i, using the sets of eight, 16, and 32 teams that team i could possibly play in rounds 4, 5, and 6, respectively.

Using these cumulative probabilities, we then constructed a 64 x 7 probability table comparable to those of West (2006, 2008, 2009), with each entry representing the probability that the team in the row would win exactly the number of games in the column (i.e., 0, 1, 2, 3, 4, 5, or 6). The probability that team i would win exactly zero games was simply the probability of team i not making it through the first round:

$$\hat{\pi}_{i0}^{g} = 1 - \hat{\pi}_{i1}^{c} \tag{6}$$

The probability that team i would win exactly one game was equal to the cumulative probability of winning in round 1, multiplied by the probability of losing in round 2:

$$\hat{\pi}_{i1}^{g} = \hat{\pi}_{i1}^{c} \times (1 - \hat{\pi}_{i2}^{c} / \hat{\pi}_{i1}^{c}) \tag{7}$$

Similarly and more generally, the probability that team i would win exactly n games ($1 \le n \le 5$) was computed as

$$\hat{\pi}_{in}^{g} = \hat{\pi}_{in}^{c} \times (1 - \hat{\pi}_{i,n+1}^{c} / \hat{\pi}_{in}^{c}) \tag{8}$$

Finally, the probability that team i would win exactly six games was the same as the cumulative probability of making it through the sixth and last round:

$$\hat{\pi}_{i6}^{g} = \hat{\pi}_{i6}^{c} \tag{9}$$

The sum of the predicted probabilities in each row of this matrix was by definition 1.00, and the sums of the predicted probabilities in each of the seven columns were by definition 32, 16, 8, 4, 2, 1, and 1, respectively. Thus, this probability table did not require the adjustments described in West (2006, 2008) to assure that the probabilities fit these known marginal constraints associated with the Tournament.[6,7] Using this probability table, we then computed the predicted (expected) number of wins by each team as a probability-weighted average, as done by West (2006, 2008, 2009):

$$E_i[WINS] = \sum_{n=0}^{6} (n \times \hat{\pi}_{in}^{g}) \tag{10}$$

_____

[6] By definition, 32 teams will win zero games in the Tournament, 16 teams will win exactly one game, eight teams will win exactly two games, etc.
[7] The entries in this probability table would be those that would have resulted from an infinite number of runs of a simulation of the Tournament, in which the probability of each team winning each possible game was computed using the binary logit model in question.

6. FURTHER COMPARISON TO WEST'S METHODOLOGY

As described above, our modeling methodology used (1) a Tournament game as the unit of observation, (2) a binary dependent variable reflecting whether the team with the higher RPI won the game, (3) binary logistic regression as the estimation method, and (4) the differences between the nitty-gritty metrics for each of the two teams in the game as predictor values. These choices differed from West (2006, 2008), which used (1) each Tournament team as the unit of measure, (2) the number of wins by that team as the dependent variable, (3) ordinal logistic regression (OLRE) as the estimation method, and (4) the "raw" values of each of his predictors as the independent variable values. In an effort to determine if the variables most related to Tournament performance would be the same whether we employed our approach or OLRE, we replicated West's approach for our data set. We computed the number of wins for every one of the 640 teams involved in the 10 Tournaments from 1999-2008, and used the raw pre-Tournament nitty-gritty values for each of these teams as the predictors. We then examined two OLRE models: one using the factors included in Model 1, and a second generated through a stepwise ordinal logistic regression process using a 0.20 significance level for entry and exit.[8]

7. RESULTS: SIGNIFICANT VARIABLES

The results for Model 1 are detailed in Table 2. As indicated in the table, only eight variables from the 32 potential predictors were included in the model. There was collinearity among some of the factors in the model, but the magnitude of the variance inflation factors (also reported in Table 2), computed using an ordinary least squares regression, were such that multicollinearity among the included predictors did not appear to be a problem.[9] Moreover, when examined as lone predictors in separate logit models, in the absence of the other factors each of the first six variables in Model 1 was significant and maintained the sign of its respective coefficient. However, the signs of the coefficients for the last two predictors in Model 1 were both the opposite of what would be expected, and were the opposite of their coefficient signs when each was examined as a lone predictor. Thus, it appears that these two factors are serving as proxies for some other omitted factor(s) with which they may be correlated. It is also possible that each of the other factors individually, or linear combinations of any subset thereof, are also serving as proxies of omitted factors.

---

[8] As in West, we omitted the play-in game loser from the data set, and did not count the play-in game as a win for the team that won that game.
[9] The complete matrix of correlation coefficients is available by request from the authors.

Table 2. Results for our stepwise binary logit model (Model 1), and with difference in seed added (Model 2).

	Model 1				Model 2			
Games classified correctly	467 (73.2%)				471 (73.8%)			
Max-rescaled R-square	0.2817				0.2863			
Hosmer-Lemeshow GOF	6.90 (p=.5470)				2.55 (p=.9594)			
AIC	660.05				659.49			
Schwarz criterion	700.17				704.07			
Variable	**Coeff.**	**p**	**Odds Ratio**	**VIF**	**Coeff.**	**p**	**Odds Ratio**	**VIF**
INTERCEPT	1.2549	<.0001			1.2478	<.0001		
DCONFWINS	0.9764	<.0001	2.66	2.53	0.8440	<.0001	2.33	3.11
DNCWINS	0.5102	0.0004	1.67	1.83	0.4228	0.0065	1.53	2.12
DNCSOS	0.3224	0.0031	1.38	1.14	0.2915	0.0086	1.34	1.18
DCRANKNC	-1.1275	<.0001	0.32	2.19	-0.9234	<.0001	0.40	3.72
DRCHAMP	0.3566	0.0089	1.43	1.80	0.3674	0.0073	1.44	1.80
DT25WINS	0.3454	0.0101	1.41	1.76	0.2980	0.0300	1.35	1.84
DT25BELOW500	0.2875	0.0117	1.33	1.51	0.2982	0.0094	1.35	1.52
DNEUTABOVE500	-0.2898	0.0165	0.75	1.48	-0.2523	0.0408	0.78	1.54
DSEED					-0.2779	0.1110	0.76	3.54

DCONFWINS: Difference in number of conference wins
DNCWINS: Difference in number of non-conference wins
DNCSOS: Difference in non-conference strength of schedule
DCRANKNC: Difference in rank of team's conference in that season, according to the mean non-conference RPI of all teams in the conference
DRCHAMP: Difference in the binary variable reflecting whether a team won its regular season conference championship (or co-championship)
DT25WINS: Difference in the number of wins vs. teams ranked 1-25 in RPI
DT25BELOW500:   Difference in the number of losses below a 0.500 record against teams ranked 1-25 in RPI
DNEUTABOVE500: Difference in number of neutral court wins above a 0.500 record
DSEED: Difference in seed

Adding the variables associated with the round in which the game was played did not change Model 1. When running a stepwise procedure in which the original eight factors in Model 1 were forced into the model, none of the five round dummies and none of the 8x5 = 40 interactions of round dummies with nitty-gritty factors were selected and significant even at the 0.10 level. Alternatively, when the original eight predictors were left open for selection along with all 45 round variables, the original eight predictors were still the only ones selected.

We also found no impact associated with forcing the difference in overall RPI, the difference in conference RPI, and the difference in non-conference RPI into a stepwise model. When overall RPI was forced in, the stepwise procedure again simply selected all the factors of Model 1 to be included in the model. The difference in overall RPI was insignificant (p=0.5168) in the presence of those eight factors, whereas the eight Model 1 factors each maintained statistical significance in the presence of the overall RPI. Similar results were generated when the difference in conference RPI or the difference in non-conference RPI were each forced in individually; conference RPI (p=0.7190) and non-conference RPI (p=0.2441) were each insignificant in the presence of the eight factors in Model 1. We also tried forcing in the conference RPI and the non-conference RPI as a group. Again, the stepwise procedure then simply selected the eight factors of Model 1, and these two RPI factors were left insignificant, with p-values of 0.9282 and 0.2664, respectively.

We did find weak significance associated with the difference in seeds when that factor was added to Model 1. Table 2 shows the results of the additional logit run of Model 1 when also including the difference in seeds in the same model; we refer to this additional estimation as Model 2. The difference in seeds was statistically significant at the 0.10 level (one-tailed) in the presence of Model 1's eight factors. However, the Model 1 factors maintained roughly the same coefficient values as well as their significance in Model 2. This result suggests that Model 1 is not simply capturing the same information as is included in the Committee's seeds. However, it also suggests that the Committee's work is generating a small degree of predictive information over and above that reflected in the nitty-gritty report.

8. RESULTS: MODEL FIT AND PREDICTIVE POWER

In terms of fit to the sample data, Model 1 correctly classified the winner in 467 of 638 games (or 73.2%). Model 2's addition of the difference in seeds to the Model 1 factors improved the number of in-sample games correct to 471 (73.8%). By comparison, the RPI alone correctly classified 443 of 637 games (or 69.6%, not counting the one game where the two teams' RPI's were identical), and the

Sagarin ranking correctly classified 452 of 638 winners (or 70.8%). Not counting 16 games where the seeds were the same for the two teams, the Committee's seeds correctly classified 441 of 622 game winners (or 70.9%), whereas Model 1 correctly classified 460 (74%) of those games. We emphasize that we mention the accuracies of the seeds and the Sagarin ranking only to put the fit of Model 1 into some perspective. A direct and formal comparison of accuracy would be inappropriate, as the Model 1 results are in-sample whereas those of the benchmarks are out-of-sample. However, the results indicate that Model 1 appears to have reasonably good fit to the data, as at least its in-sample accuracy is better than each benchmark's out-of-sample accuracy. Had the reverse been true, it would have certainly called into question the fit of the model.

In order to statistically examine model fit more formally, we performed Hosmer-Lemeshow goodness-of-fit tests for Model 1 (see again Table 2), as well as for Model 2. The test on each model indicated no significant concerns regarding fit. Negelkerke's maximum re-scaled R-square, the Akaike Information Criterion, and the Schwarz Criterion were each virtually identical for Model 1 and Model 2, again suggesting little marginal contribution by adding seed information to Model 1.

The predictive performances for the 2009 Tournament are shown for each model in Table 3.[10] Model 1 correctly predicted the winner in 47 of the 63 games in 2009 (or 74.6%). The 47 games correct in 2009 was essentially identical to the model's 46.7 average over the 10-year period (1999-2008) on which it was built. It was also better than the 46 games that Model 2 predicted correctly in 2009, indicating that the additional seed information in Model 2 actually slightly worsened predictive performance in terms of number of games correct.

Table 3 also contains the number of games predicted accurately in 2009 by the three West models, the Sagarin ranking, the RPI alone, the seeds assigned to each team, the Las Vegas betting favorites, and Kvam and Sokol's LRMC method. Model 1 performed better than the LRMC method, which predicted 45 games correctly (71.4%). Model 1 matched West's two OLRE models, the RPI, the seeds, and the betting favorites, as each of these also predicted 47 games correctly, while the Sagarin ranking and West's Bradley-Terry simulation model fared one game better at 48 games correct (76.2%).[11]

---

[10] The results shown in Table 3 do not include the 2009 play-in game, which was missed by Models 1 and 2, and correctly predicted by the RPI, Sagarin, Kvam and Sokol's LRMC, and the Vegas betting line. This was done to allow direct comparison of performance to that of the three West (2009) models. (West does not include the play-in-game in his models, and therefore also does not report metrics for either participant.)

[11] The Sagarin ranking, the RPI, and the seeds each had an unusually good year in 2009, as the expected number of correct games from each (based on historical accuracies during 1999-2008) was 44.6, 43.8, and 44.7, respectively.

Table 3. Predictive accuracy for 63 games of 2009 Tournament.

Method	Games Predicted Accurately	Sum of Squared Errors (SSE)	Upsets Predicted
Model 1	47	46.39	6
Model 2	46	46.01	6
West's OLRE (unadjusted)	47	51.85	2
West's OLRE (adjusted)	47	50.16	2
West's Bradley-Terry	48	55.04	2
RPI	47	43.74*	3
Sagarin	48	51.06*	2
Seeds	47	48.88*	N/A
Kvam and Sokol LRMC	45	N/A	3
Las Vegas betting favorites	47	N/A	1

* SSE values for the RPI, Sagarin, and seeds were computed from binary logit models using difference in the respective metric as the sole predictor.

When Model 1 was used as described earlier to generate the expected number of games won by each of the 64 teams that competed in the 2009 Tournament, the sum of squared errors (SSE) for Model 1 was 46.39. As shown in Table 3, that SSE was superior to the SSE values generated in 2009 by all three models presented in West (2006, 2008), for which the best SSE was 50.16. Again, the additional seed information in Model 2 added little to performance, as Model 2's SSE was virtually identical to that of Model 1 (46.01 vs. 46.39).

To further examine the comparison to the benchmarks, we performed additional binary logit runs using the difference in the RPI, the difference in the Sagarin ranking, and the difference in the seeds as sole predictors in respective binary logit models fit to our data. We then converted the predicted values from these three binary logit models into a predicted number of games won by each team, using our previously described procedure. In these comparisons, Model 1 again performed well. As shown in Table 3, Model 1 did worse than the difference in RPI model, but was better than the difference in the Sagarin ranking and the difference in the seeds at predicting the number of games won by each team.

Finally, as was done in West (2008), we also report in Table 4 the number of upsets (i.e., games in which the worse seed defeated the better seed) that were predicted by each approach in 2009. Model 1 predicted six of the 16 upsets in

Table 4. OLRE results using factors from Model 1, and resulting from stepwise procedure.

	OLRE using Model 1 Factors		OLRE from Stepwise		Binary logit estimation using differences in stepwise OLRE factors	
Games classified correctly	N/A		N/A		468	
Max-rescaled R-square	0.4692		0.4750		0.2620	
Hosmer-Lemeshow GOF	N/A		N/A		3.34 (p=.9109)	
AIC	1405.31		1399.10		671.05	
Schwarz criterion	1467.77		1461.56		711.17	
2009 games correct	44 / 63 = 69.8%		46 / 63 = 73.0%		47 / 63 = 74.6%	
SSE of predicted games won	62.35		53.41		39.60	
Variable	**Coeff.**	**p**	**Coeff.**	**p**	**Coeff.**	**p**
INTERCEPT 6	-5.7774	<.0001	-5.8558	<.0001		
INTERCEPT 5	-4.9731	<.0001	-5.0572	<.0001		
INTERCEPT 4	-4.0390	<.0001	-4.1237	<.0001		
INTERCEPT 3	-2.9781	<.0001	-3.0559	<.0001		
INTERCEPT 2	-1.7951	<.0001	-1.8686	<.0001		
INTERCEPT 1	-0.1545	0.1545	-0.2123	0.0592	1.2186	<.0001
CONFWINS	0.9925	<.0001				
NCWINS	0.5086	<.0001				
NCSOS	0.2123	0.0177				
CRANKNC	-1.4931	<.0001	-1.2549	<.0001	-0.8807	0.0002
RCHAMP	0.3721	0.0010	0.2377	0.0453	0.2658	0.0524
T25WINS	0.4729	0.0001				
T25BELOW500	0.1905	0.0546				
NEUTABOVE500	-0.2625	0.0049	-0.6329	0.0003	-0.6375	0.0028
RPI			0.7158	0.0016	0.5329	0.0222
ABOVE500			0.6315	0.0002	0.4811	0.0123
NEUTWIN			0.4927	0.0032	0.5378	0.0078
T25ABOVE500			0.2685	0.0038	0.1857	0.1204
T100BELOW500			-0.3498	0.0310	-0.1461	0.2885

RPI: Overall RPI
ABOVE500:   Number of conference wins above a 0.500 record
NEUTWIN: Number of neutral court wins
T25ABOVE500:   Number of wins above a 0.500 record against teams ranked 1-25
T100BELOW500: Number of losses below a 0.500 record against teams ranked 51-100

2009, a total that substantially surpassed the totals of every comparative approach in the table. The only exception was Model 2, for which the additional seed information did not improve the number of upset predictions by Model 1.

In sum, the fit and predictive results indicate that Model 1 appears to be a reasonably strong model of Tournament performance.

9. OLRE ESTIMATIONS

The first two columns of Table 4 contain the results of our two ordered logit regression estimations, using an approach comparable to that of West (2006, 2008) in which the dependent variable was the number of Tournament wins by each team during the year in question. When using this alternative methodology, the factors in Model 1 largely maintained their statistical significance. The only exception was the number of losses below break-even against top 25 opponents, for which the significance level moved to just outside 0.05. However, the stepwise OLRE approach selected a rather different set of predictors than those identified by our original stepwise procedure.

In response, we used our original binary logit approach to examine the set of predictors identified by the OLRE stepwise. The result of this estimation is shown in the last column of Table 4. As shown in the table, two factors (the last two in the table) that were shown to be significant in the OLRE model are actually not significant (at the 0.10 level, two-tailed) when using our modeling approach. This finding appears to underscore the aforementioned concern with controlling for the variation associated with the seeding decisions of the committee.

In Table 4 we have also included the 2009 predictive performances of the two OLRE models, as well as the extra binary logit model that included the set of predictors from the stepwise OLRE. The OLRE version of Model 1 generated worse predictive performance in 2009 than its corresponding binary logit. Moreover, the stepwise OLRE also generated worse 2009 performance than the corresponding binary logit using the same factors from that stepwise OLRE. Again, these findings appear to emphasize the advantage of treating the game instead of the team as the unit of measure.[12]

----

[12] Interestingly, although the in-sample fit statistics for the binary logit using the stepwise OLRE factors were not necessarily improved over those of Model 1, and despite its inclusion of two insignificant predictors, its 2009 predictive performance (in terms of SSE) was actually the best of all the models we examined, and surpassed all the comparison benchmarks. Whether this is a simply a one-season anomaly bears future research attention.

10. MODEL DISCUSSION

To the follower of college basketball, the first six variables included in Model 1 should have good face validity. The model indicates that teams with more wins before the Tournament (in and out of the conference) tend to win more often during the Tournament. Moreover, the higher the quality of the competition against whom those in-conference wins come – as represented by the conference rank variable – and against whom those out-of-conference wins come – as represented by the non-conference strength of schedule factor – the more frequently such a team wins in the Tournament. Additionally, more wins against top competition appear to be related to better performance in the Tournament. Finally, teams that were their conferences' champions over the long-haul of the regular season are more apt to win as well, when playing opponents that did not win their respective leagues.

The significance of the non-conference strength of schedule variable is particularly noteworthy, as recent Committee chairs have repeatedly pointed to that factor as one in which they place significant weight. Thus, Model 1 supports this choice by the Committee. The presence of the conference ranking factor – reflecting the differences in the ranking of the respective conferences from which the two teams come – is also interesting. The negative coefficient reflects the fact that the team from the better conference will have a conference ranking that is smaller (numerically) than its opponent's – and suggests that teams from better conferences tend to win games (all other things being equal). This finding supports a notion held by some college basketball observers – perhaps to some extent the Committee itself (Bowlsby 2005) – that teams from better-performing leagues should be favored by the Committee. This finding is particularly interesting given that the size of this factor's coefficient is easily the largest of all factors in either model.

The odds ratios shown in Table 2 speak further to the relative impact of each factor. For example, the coefficient of 0.9764 in Model 1 for the difference in the number of conference wins indicates that a one standard deviation increase in this difference will multiply the odds of the higher-ranked team in RPI winning the game by a factor of $exp(0.9764) = 2.66$, or an increase of 166%. This is substantially larger than the 67% increase associated with a one standard deviation increase in the difference in the number of non-conference wins. Both of these generate larger impact than one standard deviation changes in every other factor in the model, save for the difference in conference ranking.

The coefficient of -1.1275 in Model 1 for the difference in conference ranking indicates that a one standard deviation increase in this difference multiplies the odds of the higher RPI team winning the game by $exp(-1.1275) = 0.32$. Since higher values represent *worse* factors for this factor, the odds ratio of

0.32 is equivalent to a 1 / 0.32 = 3.13 odds ratio were it reverse-scored. This implies a 213% increase in the odds of the higher RPI team winning.

We interpret the last two factors of Model 1 with some hesitancy, as reasons for the unexpected coefficient signs for the difference in the number of games below 0.500 against top 25 opponents and number of games above 0.500 on neutral courts is certainly a matter of conjecture. It's possible that each may reflect an additional strength of schedule component over and above that captured in the other factors in Model 1. That is, teams that are further below 0.500 against top 25 teams to some extent are teams that actually *play* more top 25 teams: that is, a team can't be well below 0.500 against top 25 opponents if it didn't play any, or if it only played very few. Somewhat similarly, teams that are further above 0.500 on neutral courts may be teams that tend to schedule easier opponents in such venues.

Perhaps as noteworthy as the variables included in the model are some that aren't. Our findings show that in the presence of the factors in Model 1, the oft-discussed RPI (in any of its forms) is not significantly related to Tournament performance. However, this is not necessarily surprising, given that the factors in the model reflect various elements of the RPI (e.g., wins, opponent strength). Moreover, the RPI does show up indirectly in the model as a means of assessing groups of opponents (e.g., through the conference rank and non-conference strength of schedule variables).

However, other factors that receive considerable annual attention also do not appear in the model. Absent is the team's record in the last 10 games, which is one of the most commonly discussed performance metrics both prior to the Committee's decision announcements (including occasionally by the Committee (Bowlsby 2005; Littlepage 2006a; O'Connor 2008a)), and afterward when fans try to project game winners as they complete their pre-Tournament brackets. Teams that are "hot" coming into the Tournament do not tend to win more games once they are there, or at least the factors in Model 1 already capture this information. Thus, the attention given this factor appears unwarranted.

Other than the obvious exception of the two Model 1 factors associated with performance against top 25 teams, also noteworthy is the absence of any factor reflecting wins and/or losses against teams ranked at various positions in the RPI. For example, so-called "bad losses" against teams ranked below 100 in RPI have no more relationship to Tournament performance than any other loss. This finding is also interesting given comments by Committee chairs regarding the relevance of such factors in their decisions (Littlepage 2006a; Prisbell 2008).

Media and fans often point to those teams that received automatic bids – which are almost exclusively achieved through winning their respective conference tournaments – as more likely winners of Tournament games, due to the presumed "hotness" effect. However, this factor was also found to be

unrelated to Tournament success. Indeed, the absence of this factor combined with the presence of regular season championships in Model 1 points to an unsurprising statistical finding. That is, the relatively large sample of games over the course of an entire conference season, in which teams either play a round-robin schedule or something fairly close it, and during which teams often play a "home-and-home" format against each conference opponent, appears to yield a more accurate assessment of who the best teams are than does the small sample of games that take place in post-season conference tournaments. This is also an important managerial finding, as it raises an obvious indictment of conferences that crown their tournament winner as their NCAA Tournament representative. It also raises an indictment of stated seeding preferences given to conference tournament winners over regular season champions by the Committee (Bowlsby 2005). For that matter, it presents a critique of the NCAA Tournament itself, which arguably suffers from the same shortcomings associated with small sample sizes.

11. CONCLUSION

It appears that the Committee's nitty-gritty reports do contain information that is quite representative of team strengths. Thus, from that perspective, the reports appear to be useful input for the Committee's seeding processes. However, our findings suggest that of the myriad of data made available on these reports, only eight factors appear most relevant in assessing team strengths. To the extent that the Committee wishes the reports to reflect information that is strictly relevant to the Tournament performance of teams once they are in the Tournament, our findings imply that the Committee could substantially reduce the amount of data included on the report without loss of information. Although most of these factors have good face validity, they are not necessarily the ones that are commonly highlighted either in the statements of the Committee or in the discussions from media, fans, and participants.

REFERENCES

Bowlsby, B. (2005), "NCAA Press Conference, March 13, 2005," www.asapsports.com/basketball/2005college/031305BB.html (accessed June 15, 2005).

Boulier, B. L. & Stekler, H. O. (1999), "Are Sports Seedings Good Predictors?: An Evaluation," *International Journal of Forecasting*, 15, 83-91.

Boulier, B. L. & Stekler, H. O. (2003), "Predicting the Outcomes of National Football League Games," *International Journal of Forecasting*, 19, 257-270.

Brown, G. T. (1999), "Dissecting the Rating-Percentage Index,"
http://www.ncaa.org/wps/ncaa?ContentID=32948 (accessed April 29,
2009).

Carlin, B. P. (1996), "Improved NCAA Basketball Tournament Modeling Via
Point Spread and Team Strength Information," *The American Statistician*,
50 (1), 39-43.

Caudill, S. B. (2003), "Predicting Discrete Outcomes with the Maximum Score
Estimator: The Case of the NCAA Men's Basketball Tournament,"
International Journal of Forecasting, 19, 313-317.

Cohen, Y. (2006), "New RPI Formula Has Helped Mid-Majors."
msn.foxsports.com/cbk/story/5388240 (accessed June 29, 2006).

Coleman, B. J. (2008), "NCAA Men's Basketball Tournament Dance Card,"
DanceCard.unf.edu (accessed March 24, 2008).

Coleman, B. J. & Lynch, A. K. (2001), "Identifying the NCAA Tournament
'Dance Card,' " *Interfaces*, 31 (3), 76-86.

Covers.com (2009), http://scores.covers.com/college-basketball-scores-
matchups.aspx?t=0 (accessed April 10, 2009).

Harville, D. A. (2003), "The Selection or Seeding of College Basketball or
Football Teams for Postseason Competition," *Journal of the American
Statistical Association*, 98 (461), 17-27.

Henderson, J. (2008), "Bracketology 101: Mastering Method to March Madness
Maniacal," www.denverpost.com/sports/ci_8213646 (accessed Sep 24,
2008).

Hosmer, D. W. and Lemeshow, S. (1989), *Applied Logistic Regression,* New
York: John Wiley & Sons, Inc.

Jing, Y., Cox, A. (2008), "Crashing the Dance: Learning the Behavior of the
NCAA Basketball Selection Committee,"
www.cc.gatech.edu/~yjing/work/cs6230_project1.pdf (accessed Jan 14,
2008).

Kvam, P. & Sokol, J. S. (2006), "A Logistic Regression / Markov Chain Model
for NCAA Basketball," *Naval Research Logistics*, 53 (8), 788-803.

Littlepage, C. (2006a), "National College Athletic Association Media Conference
(February 22)," www.asapsports.com/show_conference.php?id=38807
(accessed May 8, 2008).

Littlepage, C. (2006b), "An Interview with Craig Littlepage (March 12),"
www.ncaa.com/basketball-mens/article.aspx?id=2772 (accessed April 6,
2008).

National Collegiate Athletic Association (2005a), *Division I Men's Basketball
Championship Handbook*, 68,
www.ncaa.org/library/handbooks/basketball/2005/2005_d1_m_basketball
_handbook.pdf (accessed June 7, 2005).

National Collegiate Athletic Association (2005b), "NCAA Division I Men's Basketball Championship Principles and Procedures for Establishing the Bracket (November 9, 2005)," www.ncaasports.com/basketball/mens/story/6985142 (accessed Jan 18, 2008).

National Collegiate Athletic Association (2006a), "NCAA Division I Men's Basketball Championship Principles and Procedures for Establishing the Bracket (7/01/06)," www.ncaa.org/wps/wcm/connect/resources/file/ebada14a14711d0/Principl es_And_Procedures.doc?MOD=AJPERES (accessed May 28, 2008).

National Collegiate Athletic Association (2006b), "NCAA Division I Women's and Men's Basketball Committees Announce Regular Release of Ratings Percentage Indexes," www2.ncaa.org/portal/media_and_events/press_room/2006/january/20060 123_rpi_rls.html (accessed Jan 18, 2008).

National Collegiate Athletic Association (2007), "NCAA Division I Men's Basketball Championship Principles and Procedures for Establishing the Bracket (7/31/07)," www.ncaa.org/wps/wcm/connect/resources/file/ebac4d4a13e0063/principl esandprocedures_2008.pdf?MOD=AJPERES (accessed Jan 14, 2008).

O'Connor, T. (2008a), "Tom O'Connor Pre-Selection Teleconference (March 12)," www.ncaa.com/basketball-mens/article.aspx?id=131242 (accessed April 4, 2008).

O'Connor, T. (2008b), "An Interview with: Tom O'Connor (March 16)," www.ncaa.com/basketball-mens/article.aspx?id=142040/ (accessed May 8, 2008).

Palm, J. (2009), www.collegerpi.com (accessed April 20, 2009).

Prisbell, E. (2008), "Committee's Decisions Came Down to Final Buzzer," *Washington Post*, March 17, www.washingtonpost.com/wp-yn/content/story/2008/03/17/ST2008031700082.html (accessed April 4, 2008).

Sanders, S. (2007), "A Cheap Ticket to the Dance: Systematic Bias in College Basketball's Ratings Percentage Index," *Economics Bulletin*, 4 (34), 1-7.

SAS Version 9.1, (2004), SAS Institute, Inc., Cary, North Carolina.

Schwertman, N. C., McCready, T. S., & Howard, L. (1991), "Probability Models for the NCAA Regional Basketball Tournaments," *The American Statistician*, 45 (1), 35-38.

Schwertman, N. C., Schenk, K. L., & Holbrook, B. C. (1996), "More Probability Models for the NCAA Regional Basketball Tournaments," *The American Statistician*, 50 (1), 34-38.

Smith, T. & Schwertman, N. C. (1999), "Can the NCAA Basketball Tournament Seeding be Used to Predict Margin of Victory?" *The American Statistician*, 53 (2), 94-98.

Sokol, J. & Nemhauser, G. (2009a), "Pure LRMC Rankings through games of 3/15/2009." http://www2.isye.gatech.edu/~jsokol/lrmc/ (accessed April 13, 2009).

Sokol, J. & Nemhauser, G. (2009b), "The LRMC Tool for NCAA Tournament Selection and Seeding." http://www2.isye.gatech.edu/~jsokol/lrmc/lrmcsummary.pdf (accessed April 13, 2009).

Stern, H. S., Massey, K., Harville, D. A., Billingsley, R., Colley, W., & Mease, D. (2004), "Statistics and the College Football Championship / Discussion / Reply," *The American Statistician*, 58 (3), 179-195.

Stern, H. S. & Mock, B. (1998), "College Basketball Upsets: Will a 16-Seed Ever Beat a 1-Seed?" in the column "A Statistician Reads the Sports Pages," *Chance*, 11 (1), 26-31.

West, B. T. (2006), "A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament," *Journal of Quantitative Analysis in Sports*, 2 (3), Article 3.

West, B. T. (2008), "A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament: Updated Results from 2007," *Journal of Quantitative Analysis in Sports*: 4 (2), Article 8.

West, B. T. (2009), http://www-personal.umich.edu/~bwest/new_ratings_2009.xls (accessed April 13, 2009).

Zuchowski, M. (2005), "Mock NCAA Selection Committee, Journal #2." marquette.scout.com/2/353964.html (accessed June 29, 2006).