

# **Evaluating Regional Balance in the NCAA Men's Basketball Tournament using the Tournament Selection Ratio**

John A. Trono  
Saint Michael's College  
Colchester, VT 05439 (USA)  
jtrono@smcvt.edu

## **Abstract**

The placement of teams, or players, into a tournament's brackets can have a significant impact on who will emerge as the champion. At Wimbledon, 32 players are seeded among the 128 singles entrants in that particular tennis tournament. UEFA coefficients are used to select those European soccer clubs that will compete for the Champions League and Europa League titles in the following season. Teams that survive qualifying are seeded into groups for the finals of World Cup competition in soccer, basketball and ice hockey. An important question to answer is whether or not the criteria used to determine such seeds will create an equitable tournament bracket. The Tournament Selection Ratio (TSR) was designed specifically to quantitatively address this question with regards to the annual NCAA men's basketball tournament, where 64 teams are selected from over 340 teams located across the entire United States. The TSR metric will be used to evaluate: if the four regions (16 seeded teams in each) in recent NCAA Men's basketball tournament have been evenly balanced; if the appointed selection committee, who creates the tournament bracket, has invited the best teams to compete in this sport's culminating event; and if those invited teams have been assigned the most accurate seeds. The results of current NCAA practices and procedures, as applied by the selection committee for this tournament, will also be compared against previous invitational guidelines.

## **Introduction**

The National Collegiate Athletic Association (NCAA) is the official body that oversees every intercollegiate sport in the United States. In almost every one of these sports, a championship tournament is held – after the regular season, and after all postseason, conference tournaments (in most sports) have finished – to allow direct competition to determine who will have earned the annual, prestigious title of national champion. With regards to basketball, the NCAA men's tournament is a roughly month long event where every eligible team has the opportunity to earn an invitation to participate in this extravaganza.

Any team that can win its conference's postseason tournament is automatically invited, and almost half of the teams competing in the NCAA basketball tournament receive these automatic bids; the remaining tournaments spots are filled by the worthiest teams as chosen by the NCAA's selection committee (for this tournament). Starting in 1985, 64 teams have battled to become the national champion in this tournament, with 16 teams (seed #1 to #16) competing in four separate regions. The four regional champions continue on to what is known as the Final Four, where three more games will produce the winner of this single elimination tournament; that team, who has outlasted the entire field of 64, then receives the national champion's trophy.

The number of conferences receiving automatic bids has varied between 29 and 32 since 1985, and 31 tournament positions have been allocated each year since 2001. In 2001, the field was also expanded to 65 teams, with the two teams from the weakest conferences (that typically have received automatic bids) competing in a 'play-in' game. The field was expanded again to 68 teams in 2011, now with four play-in games. These additional invitations have increased the selection committee's chances of not excluding qualified teams when selecting who will receive the remaining at-large bids. However, there is always plenty of discussion concerning did the committee invite the best teams, from those remaining, after the automatic bids have been awarded?

The rest of this paper will describe: the Tournament Selection Ratio (TSR), and its performance predicting the teams that will be invited to the NCAA tournament by said tournament selection committee as well as how accurate the assigned seeds have been; the model used to determine the probability that a team will reach the Final Four; the TSR strength function and its predicted effectiveness regarding the teams reaching the Elite Eight (the teams still left in this tournament in the round before the Final Four begins); and finally, how balanced the four NCAA tournament regions have been since 1985 in comparison to the regional balance before the field expanded to 64 teams that year.

### **Essentials of the Tournament Selection Ratio**

The NCAA does not hold a tournament to determine who the national champion is with regards to the sport of football. Instead, the results of all the games that season inform the Bowl Championship Series (BCS) methodology, which has evolved since its inception in 1998. The TSR mimics the BCS formula in several ways. Normalized results from the two polls (sports writers in one, and coaches in the other) comprise two thirds of the BCS formula, while six computer ranking models contribute the other third. Using the trimmed Borda method, the lowest and highest computer rankings are dropped, and the other four values are averaged (and normalized) before being added into the final BCS ranking. The top two teams in the final BCS standings compete for college football's national championship.

The BCS methodology was modified in 2002 to only employ ranking models exclusively after the NCAA deciding to have all computer systems ignore margin of victory, to lessen the motivation for teams to run up large wins over weaker teams (in hopes that the computer models would consider their team as one of the best teams that year). The TSR does not solely rely on ranking models, which limits the winning point differential in any contest to be at most one point, because limiting margin of victory has been proven to decrease prediction models when forecasting future outcomes (Berry, 2003). The TSR consists of four rating and four ranking systems, utilizing the trimmed Borda method, with this average rank contributing 50% to the final TSR. The two polls, once normalized by the maximum voting points a team can receive, each add another 25% to this ratio (which ranges from zero up to a maximum of one, since the trimmed Borda mean is normalized as well). Therefore, the TSR weights equally the objective results, provided by those eight quantitative systems, with the subjective expertise of those selected to participate in the two major polls. Of the eight systems, four have been widely recognized as valuable, quantitative measures of team performance; the other four (two rating and two ranking systems apiece) have been devised by the author over the past twenty years. These eight systems will now be briefly described; more detailed explanations can be found in the references.

The Rating Percentage Index (RPI) was been a mainstay of the NCAA men's basketball tournament selection committee since the early 1980s, helping them to objectively select (and seed) the teams that have best demonstrated their worthiness to compete for this title. The RPI is also a weighted formula, like

the TSR, with a team's won-loss percentage contributing 25% to the aggregate. Another 50% comes from the average won-loss percentage of each team's opponents, with the last 25% being derived from each team's opponents' opponent's won-loss percentage. (This formula was modified in 2004 to weight road wins more than wins earned at a neutral site, which are weighted more highly than wins in front of one's home fans. All references below are with respect to the original RPI formula, which I refer to as RP.)

The power rating system (PW, as described in Carroll, et al, 1988) has been verified to be a very accurate predictor of games not yet played (Trono, 2010). This system iterates over all games that season until each team's strength of schedule component, which directly impacts the overall rating a team is assigned when applying this technique, stabilizes to within a specified tolerance. If the scores are modified so that the largest margin of victory is at most one, then this rating system becomes a ranking system (P1).

Another ranking system, one that was designed for possible inclusion in the BCS formula, is the Rewards system (Trono, 2007). In this system (RW), an average win value is computed by weighting each win in an exponential fashion, once those wins are sorted from best to worst (using the aforementioned P1 system). The team appearing at top of the P1 ranking will have a rating of roughly one, and the rating for the team at the bottom will be roughly zero, with the rest of the team's ratings being normalized in accordance to their P1 rating (versus the best and worst teams). The normalized P1 rating of the best opponent a team defeats is weighted roughly 40% more than the second strongest beaten opponent, which is roughly 40% more than the third best, and so on. Each team's average win value is then multiplied by the number of wins by that team that year, and losses lessen this amount, which produces the team's final Rewards rating, generating the ranking according to this system.

The fourth ranking system included in the TSR (MD, the modified percentage stabilizer) re-computes a team's "winning percentage" in an iterative manner as well to determine how much a win over an opponent should contribute to its modified percentage. Starting with the actual percentage, a new percentage is computed by adding the opponent's percentage in the previous iteration for each win, and subtracting  $(1 - \text{opponent's percentage})$  for each loss. Then, each team's new percentage is just this accumulated sum divided by the total number of games they played. After a normalization step, once every game has been considered, this process continues until the new set of percentages converges to be within a specified tolerance of the percentages determined in the previous iteration. Therefore, wins over weaker teams will probably lower a team's next computed percentage, and losses to strong teams only slightly penalize the losing team.

The Sagarin (SG) ratings, for every NCAA football and basketball team, have been included in the *USA Today* daily newspaper since the mid-1980s. Jeff Sagarin devised his rating system to be an undisclosed combination of his pure rating system (that appears to have strong similarities to PW), with his pure ranking system (that is modeled after the ELO chess rating system). With regards to recent NCAA men's basketball tournament games, the opening Las Vegas betting lines have closely matched the differences between the competing team's Sagarin ratings.

The PW system is the sum of a team's average score differential (offensive average – defensive average) and its computed strength of schedule. The expected difference (ED) system relies solely on the average score differential to generate each team's rating. If team A has an expected score of X  $((A's \text{ offensive average} + B's \text{ defensive average})/2)$ , and its opponent's expected score is Y  $((B's \text{ offensive average} + A's \text{ defensive average})/2)$ , then the actual game differential above (or below) X-Y is added to A's rating, and likewise with Y-X and B's rating. Good teams typically perform better than the expected

game score; therefore, their ratings increase, especially when playing more games against strong teams than weaker ones.

The eighth system is a modification of the discrete rating system (DIS) which is essentially integer-based, as opposed to the seven previously described systems – all of which are based on continuous mathematics. In the DIS (Trono, 2010), each team has an integer rating, and this is multiplied by a specified, constant point value; the difference between the two team’s ratings can therefore be used to generate a predicted point spread for a game between them. If that game’s actual point spread differential is within a certain range, i.e. plus or minus a specified threshold, then there is no need to update either team’s rating. However, if the favorite (let’s say team A) wins by more than the predicted differential plus the threshold, its rating is incremented and its opponent’s (team B) rating is decremented; likewise, if A loses and/or doesn’t cover the predicted spread minus the threshold, its rating is decremented and B’s is incremented.

All teams start the season with a rating of zero, and the ratings earned by the end of the season then become a team’s initial rating for the next time that same season’s games are used to update the ratings; all team’s ratings are recomputed in this fashion until every set of ratings matches a set of ratings that have been previously generated, most likely the previous iteration – though some cycles do appear as this stabilization process unfolds. Many teams will finalize to the same integer rating, so to break these ties, penalties for losses are then applied to the integer rating (times the constant point value) almost guaranteeing unique teams ratings for the stabilized DIS (SD). (As reported in Trono, 2010, a threshold value of 1, and a rating point equaling 1 as well, had the best retrodictive prediction accuracy of all combinations when examining performance after the season is over.)

The eight systems used to compute the TSR were chosen to include a wide variety of approaches, and the ratings/rankings were either easy to obtain (SG), or readily computable by software implementation (the other seven). A breakdown of the top ten teams in 2011, according to the TSR, is provided in Table 1. (AP and COA represent the normalized poll quantities, sports writers and coaches respectively; the specific values listed beneath the 8 system names, will be explained more fully a little later on.)

#	TSR	W	L	AP	COA	PW	P1	RP	RW	EX	MD	SD	SG	Team Name
1	0.99385	32	2	0.991	0.994	70	70	69	70	69	69	70	70	OhioSt
2	0.97900	32	2	0.969	0.966	68	69	70	69	70	70	69	69	Kansas
3	0.93419	30	4	0.906	0.907	69	67	67	66	68	67	67	68	Duke
4	0.89987	27	5	0.865	0.858	66	66	63	67	65	65	66	66	Pittsburgh
5	0.87925	32	2	0.814	0.799	56	68	68	68	61	68	68	67	SanDiegoSt
6	0.84121	26	6	0.820	0.783	60	65	62	64	57	64	57	63	NotreDame
7	0.79859	26	7	0.732	0.748	57	62	64	60	58	63	60	56	NorthCarolina
8	0.76553	27	7	0.658	0.609	67	58	59	62	67	60	65	64	Texas
9	0.75177	30	4	0.601	0.568	58	64	66	65	64	66	62	65	BYU
10	0.72076	26	9	0.627	0.646	55	61	60	59	48	59	47	57	Connecticut

Table 1 – Full breakdown for top 10 teams in 2011 (according to the TSR)

### TSR’s Invitation Performance

Now that the basic elements composing the TSR have been described, it is appropriate to evaluate its efficacy. A model known simply as the “Dance Card” (DC) has been developed (Coleman and Lynch,

2001); it is a statistical model that was intended to capture the behavior of the NCAA selection committee, regarding the teams they have invited to the men’s basketball tournament. The Dance Card was trained using previous NCAA tournaments (1994-1999), and its performance has been measured ever since. Table 2 illustrates that the TSR has averaged over 90% accuracy when predicting what teams will receive the at large tournament invitations, which is only slightly less than the DC model – and the TSR had no training whatsoever!

The DC model incorporates the RPI, team records against top 50 and top 100 teams, inter-conference records, post season tournament performance, etc., so there is some overlap with what is employed by the TSR. However, the TSR does seem to capture a team’s merits, as witnessed by the strong correlation with the selection committee. (More specifics concerning years when certain teams may have been “wrongly” excluded from the tournament can be found in one of the appendices in the lengthy, comprehensive report: Trono, 2013.) The ranking produced via the TSR also correlates nicely with the teams’ seeds in this tournament, as decided by the selection committee. Considering only the invited teams, and assigning the teams with the four highest TSR values a #1 seed, and the next four as #2 seeds, and so on, the average seed difference (between what TSR would have assigned, versus the selection committee) was only 1.056: on average, for the NCAA tournaments from 1985-2011, 24 teams received the same seed in both; 24 more teams had a plus or minus one difference between the two seeding methods; 9 were plus or minus two; and 7 (out of the 64 invited teams) had a seeding difference greater than two.

This examination validates that the TRS formula devised does indeed produce a quantitative measure that matches what the qualified experts have also observed, with regards to the set of teams under scrutiny for tournament inclusion. Table 2 lists how effect each objective system has been, as compared to the DC.

Span	TSR	PW	P1	RP	RW	EX	MD	SD	SG	DC	All
01-11	340	307	346	338	324	302	332	288	337	351	377
94-00	218	185	215	212	207	187	216	179	205	224	240
85-93	277	251	275	270	260	238	269	230	265	-----	306
Total	835	743	836	820	791	727	817	697	807	575	923
Pct.	91	81	91	89	86	79	89	76	88	93	100

Table 2 - TSR performance predicting NCAA tournament at-large bids.

### Probability Models

Many models have been published that estimate the probability that teams will advance in the NCAA tournament (Brown and Sokol, 2010, and Brady, 2008). These employ quantitative measurements of each team; however, a very simple approach was proposed (Breiter and Carlin, 1997) that only relied on a team’s seed to accurately compute how likely it was that a team would reach the Final Four. Their formula for the probability that a team with seed n would defeat a team with seed k is  $k / (k+n)$ . Using this formula, they then calculated the probability that each seed would reach the Final Four. This strategy was modified (Berry, 2000) to use a seed’s strength, all sixteen of which were empirically determined to minimize the sum of the squared error terms when compared against the observed results in the NCAA tournaments from 1985 to 2000. Berry’s formula to determine the probability that seed n defeats seed k is

$\text{Strength}_n / (\text{Strength}_n + \text{Strength}_k)$ , and those seed strengths have performed just as well in all of the years after the specified training period.

To incorporate Berry's strategy, a derived TSR strength value (for each team) was required; the ordering generated by the TSR is converted into a value from 5000 (for the #1 team) down to 101 (for the 70<sup>th</sup> team in the TSR ranking). As can be seen in Table 1, the highest value listed, beneath the eight computer system names, is 70, and those values decrease indicating where that team was placed by that system, with 70 being at the top of that system's ordering. (Regarding the trimmed Borda value in the TSR, this average is divided by 70, before being multiplied by 0.5, when the TSR rating is calculated.) Because both polls included in the TSR typically only include 35 to 50 teams (each of which receives at least one vote), it seemed reasonable not to rank order all teams (from N down to 1) according to the eight objective systems. Seventy seemed like a good choice because it was only slightly larger than the number of teams invited (64), so most teams competing in the NCAA tournament should appear in some system's top 70 teams. It was also chosen because  $70^2 + 100$  is a nice 'round value'. (One hundred is added to reduce the ratio between the strength value of the top team and the #70 team when using the strength value formula:  $(71 - \text{TSR rank})^2 + 100$ . This ratio is roughly 50 (5000/101), whereas it would 4900 if that constant (100) wasn't added in the strength value formula. Since no #16 seed has ever defeated a #1 seed, a 2% chance of them winning seemed acceptable. Before 2012, only four #15 seeds have defeated a #2 seed, out of the 108 games played since 1985, so that percentage seemed to be in line with actual results, and is only slightly larger than the #1 vs. #16 percentage. Other similar ratio comparisons were also favorable – with regards to this strength formula.)

Using the updated formula, where the probability that team n defeats team k is  $\text{TSRstrength}_n / (\text{TSRstrength}_n + \text{TSRstrength}_k)$ , the probability of each team reaching the Final Four can be calculated (Berry, 2000). From 1985-2011, #1 seeds have had, on average, a 28% chance to reach the Final Four, +/- 3%, using this TSR-based model, and the #2 seeds had a 23% chance (+/- 3% as well). Teams with a high TSR ranking, who are also in a weak region, will have a higher expected likelihood of reaching the Final Four than this average.

Table 3 compares the likelihood of each seed making its way through the first four rounds (ignoring all 'play-in' games when more than 64 teams are invited) for the actual TSR strength values from 1985-2011 versus the strength values that would be assigned if the top 64 teams were invited, and the first four were #1 seeds, the next four #2 seeds, and so on down to the last four invited being #16 seeds.

	Actual Avg.	Prob.	Top 64 Avg.	Prob.
1	4768	28.10	4793	26.69
2	4185	22.94	4261	21.71
3	3678	17.26	3761	16.75
4	3212	12.60	3293	12.26
5	2679	7.27	2857	8.59
6	2310	5.08	2453	5.80
7	1831	2.60	2081	3.66
8	1655	1.85	1741	2.14
9	1367	1.03	1433	1.21
10	1189	0.71	1157	0.66

11	892	0.31	913	0.33
12	829	0.26	701	0.15
13	348	0.02	521	0.06
14	209	0.00	373	0.01
15	59	0.00	257	0.01
16	21	0.00	173	0.00

Table 3 – TSR related probability of teams to reach the Final Four.

Table 4 compares the number of teams predicted to reach the Elite Eight versus the actual counts. (The Elite Eight round has twice as many teams still competing in the tournament as the Final Four round, and so the former provides more non-zero counts for the lower seeds – for comparison purposes in Table 4.) The seed strengths, as determined by Berry empirically, are listed as well, but as can be easily observed, these values are quite uneven in their distribution whereas the TSR strength values are smoother when transitioning from the high seeds down to the lower seeds. The expected counts decrease more regularly when the TSR strength values are applied than when utilizing Berry’s trained, seed strength values.

	Actual	Top 64	TSR Str.	Berry	B (Str)
1	78	49	52	76	100
2	53	43	45	51	43
3	25	35	36	24	25
4	14	27	28	14	25
5	7	20	18	9	21
6	12	15	13	15	21
7	6	10	8	8	17
8	7	7	6	4	17
9	1	4	4	4	17
10	7	3	3	5	14
11	5	2	2	3	10
12	1	1	1	2	10
13	0	0	0	0	6
14	0	0	0	1	5
15	0	0	0	0	2
16	0	0	0	0	1

Table 4 - Expected number of teams to reach the Elite Eight: 1985-2011

### Estimating Regional Balance: 1985-2011

Given how well the untrained TSR strength values generated predictions, matching fairly closely those observed totals in Table 4, it appears that the TSR: closely matches where the selection committee places teams in the NCAA tournament bracket, and, the corresponding quadratic, TSR strength value formula

succeeds reasonably well in matching the prediction of how far teams will advance in this tournament. The TSR strength value can also be used to evaluate regional balance in the following manner. By summing the TSR strength values of all teams in a region, this will produce a quantitative measure of the relative, overall quality of teams placed into each region. Using a worst case analysis, assigning the #1, #5, #9 ... and #61 teams into one region (i.e. the best #1, the best #2, etc.) and the #4, #8, ... #64 team into another region (i.e. the worst #1, the worst #2, and so on), this produces an acceptable variation in what the TSR strength value sums could be – per region. (After performing said associated calculations, this acceptable regional sum variation is 3696.)

Table 5 lists the TSR strength value sums from 2000, along with the maximum difference between any two regions for each year. Nine of these twelve years have acceptable variations between the regions, i.e. 4300 or less, and the largest difference (7095) occurred in 2007. After careful review, that difference shrinks to 1927 if the two #6 seeds in the specific regions (East and Midwest) are interchanged. The average sum per region in Table 5 is 28980.5, and the average difference between each region’s sum, and that value, is 1293.4, which is roughly equal to one quarter of the acceptable variation that would constitute an equitable allocation of the invited teams across the regions.

Year	East	South	Midwest	West	Avg. Diff.	Max. Diff.
2000	31851	29479	25723	29025	29019.50	6128
2001	27173	31423	30173	28397	29291.50	4250
2002	26184	29563	28230	29084	28265.25	3379
2003	27974	27983	29820	31475	29313.00	3501
2004	30445	31552	25856	28882	29183.75	5696
2005	27187	30787	29894	30053	29480.25	3600
2006	27611	30229	28793	28197	28707.50	2618
2007	24103	28531	31198	29282	28278.50	7095
2008	29384	26585	29413	30057	28859.75	3472
2009	28186	29930	26906	29231	28563.25	3024
2010	28678	29306	28762	30748	29373.50	2070
2011	31212	28797	29921	27790	29430.00	3422

Table 5 – Aggregate TSR strength values (across the regions)

For the entire twenty seven year period (1985-2011), over half of those years (fourteen to be exact) the maximum, regional TSR sum difference was than 4000, with another eight years between 4000 and 6000, four more (including 2007) were between 6000 and 8000, and 1988 was the largest difference (at 11095). Even with the latter, swapping two or three judiciously chosen, same-seeded teams (between the regions with the smallest and largest TSR sums) brings even **that** bracket back into line.

Now certainly any regional imbalance can be corrected by exchanging teams between the outlying regions in question. However, when such exchanges are with teams who have been recognized as being roughly equivalent by the tournament selection committee, since those teams have been assigned the



same regional seeds, this implies that the correction is fairly minor – especially when compared to how teams were placed into regions before seeds were assigned (as described in the next section).

### Regional Balance Before 1985

The previous sections illustrated that since 1985, the NCAA tournament selection committee has been quite equitable when placing teams across the four regions, maintaining a reasonably balanced bracket in almost all cases. However, the current methodology in place has gone through quite an evolution since the first NCAA tournament in 1939. From 1951 until 1974, a tournament bracket was published before the season even began, placing designated conference champions into specified slots in geographically determined regions. This template format, known as the ‘tournament draw’, had openings for at large teams that were independent of any conference; six of the sixteen bracket slots in 1951 were reserved for these at large invitations. Prior to 1974, a serious invitation restriction was in effect then: only one team per conference could be invited to compete in this tournament. Many of the conferences automatically rewarded the team who won their conference’s postseason tournament with that conference’s automatic place in the NCAA tournament, which allowed for some really strong teams to be excluded from the NCAA tournament (when they did not win their own conference’s post season tournament).

It seems obvious that predetermining each team’s placement in the bracket (without any consideration to a team’s qualities) could be quite arbitrary, and possibly even unfair to certain teams (or regions). Teams were locked into specific regions by geographical location and/or conference affiliation, and so the top teams might meet before the Final Four, since teams were not seeded when applying the tournament draw, bracket strategy.

From 1953 to 1974, 22 to 25 teams competed in the NCAA men’s basketball tournament, with four to seven teams per region (and even that quantity varied from year to year as well). Given these circumstances, the sum including every team’s TSR strength value could be misleading, so two other values will be examined to illustrate how unbalanced the regions were before seeds were introduced into this tournament’s bracket formation in 1979.

	East	Mideast	Midwest	West
Minimum	10350	10806	5836	7385
Maximum	17598	17485	15851	17094
< 10,000	0	0	3	5
10,000-12,999	7	7	10	10
13,000-14,999	5	9	6	4
15,000-16,999	7	4	3	2
> 17,000	3	3	0	1
Average	14414.2	14245.6	12428.9	11993.4
Top four ratio	32.51%	33.50%	37.20%	39.34%
# > 40%	1	2	6	11

Table 6 – Sum of the TSR strength values for the top four teams in each region: 1953-1974

The first metric is simply the sum of the four largest TSR strength values associated with teams in each region (since four is the fewest number of teams in any region during this time period). Table 6 contains the results when applying this metric to the 22 years in question: from 1953 to 1974 (since only 16 teams were invited in 1951 and 1952). One obvious observation is that two regions (the Midwest and West) appear to be weaker than the other two during the tournament draw era. (The top four ratio in Table 6 is simply the largest TSR strength value, in each region, divided by the sum of that region's four largest TSR strengths values.) However, Table 6 would not expose weak regions if several very strong teams were to confuse this metric – when all the other teams in that region were weak ones.

Therefore, the second metric is simply the ratio of the highest TSR strength value in a region over the second highest one. Table 7 highlights (once again) that when compared to previous practices, the current methodology in place to select teams, and place them into different regions, is vastly superior to prior procedures. Only two times after 1984 has this ratio been greater than 1.3, yet in the Midwest and West regions, where apparently there are not as many strong teams/conferences as the other two, this ratio is was larger than 1.3 seventeen times.

	East	Mideast	Midwest	West
1985-2011 (Avg)	1.1502	1.1104	1.1436	1.1198
1953-74 (Avg)	1.1752	1.1152	1.2798	2.3769
>1.35	4	2	6	11
>1.5	1	1	2	6
>2.0	0	0	1	2

Table 7 – Region ratios: teams with highest TSR strength / second highest TSR strength

The final table (#8) is a composite snapshot that also magnifies how the tournament draw created mis-scheduling decisions once teams were placed into their designated, geographically aligned bracket slots. First round games were scheduled only when more than 32 teams were invited to this tournament, and each subsequent round has half the number of teams competing in the tournament as the previous round. By the fourth round, only eight teams are left, and the fifth round features the opening games in the Final Four. Table 8 lists the number of games where teams were scheduled to play **one round** before a correctly seeded tournament bracket would pair them against each other. (Table 8 does **not** include the eight games where contests were played two rounds too early.)

Span	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>
1951-74	-----	3.01	15.10	10.42	8.33
1975-78	-----	17.19	6.25	0.39	0.00
1979-84	0.00	6.25	2.50	0.00	0.00
2001-2011	5.21	3.70	5.56	0.93	0.00

Table 8 - Percentage of games where tournament pairings were one round too early

Such premature matchups would not occur at Wimbledon, for example, where a properly seeded bracket is employed. When examining such a canonical bracket of 32 players, the #1 player would be scheduled to play the player seeded as #32 in the opening round, and that winner would play the victor of the match between the #16 and #17 seeds. A match between the #3 seed and the #8 seed should not occur until the third or fourth rounds, yet such contests did occur several times, during the early transition period (1975-78) when the NCAA tournament selection process was still evolving, in the opening round!

This NCAA tournament was expanded to include 32 teams for four years (1975-78). The maximum number of teams allowed from each and every conference was also increased to two in 1975, and teams receiving at large invitations could be placed into any region. (A significant increase in the number of premature matchups occurred in this brief period than before – or after – this four year span.) The tournament held in 1979 was the first one where teams were assigned seeds, in an effort to balance the four regions; the cap, pertaining to the number of invitations allowed per conference, was also completely abolished that year. There were 40 teams invited in 1979, and this increased to 48, then 52 and eventually 64 (in 1985).

### **Summary**

The Tournament Selection Ratio (TSR) was introduced as one possible, quantitative method to evaluate how well a team has performed over the entire season. The TSR was then used to determine if the most deserving teams have been invited to the NCAA men's basketball tournament as well as generating individual strength values (per team) that can be used to predict the likelihood of a game's outcome in this tournament. Continuing such probabilistic analysis, these calculations can be extended to cover several rounds of this tournament to see how accurate these predictions/probabilities are.

In both cases, the TSR was seen to be a very accurate predictor regarding which teams would be invited to this tournament as well as deriving the expected number of rounds each team would advance in the tournament. Several other measurements (which rely only on the TSR) were taken, and these illustrated how much more evenly balanced the four tournament regions have been since 1985, when compared to the previous tournament team selection strategies that were employed (prior to 1985).

### **References**

- Berry S. (2000) A Statistician Reads the Sports Page. *Chance* 13(3), 56-61.
- Berry S. (2003) College Football Rankings: The BCS and the CLT. *Chance* 16(2), 46-49.
- Breiter, D.J. and Carlin, B.P. (1997) How to Play the Office Pools if You Must. *Chance* 10(1), 5-11.
- Brown, M. and Sokol, J. (2010) An Improved LRMC Method for NCAA Basketball Prediction. *Journal of Quantitative Analysis in Sports* 6(3), article 4.
- Carroll, B., Palmer, P. and Thorn, J. (1988) *The Hidden Game of Football*. Warner Press.
- Coleman, B.J. and Lynch, A.K. (2001) Identifying the NCAA Tournament 'Dance Card'. *Interfaces* 31(3), 76-86.
- Trono J. (2007) An Effective Nonlinear Rewards-Based Ranking System. *Journal of Quantitative Analysis in Sports* 3(2), article 3.
- Trono J. (2010) Rating/Ranking Systems, Post-Season Bowl Games, and 'The Spread'. *Journal of Quantitative Analysis in Sports* 6(3), article 6.

Trono J. (2013) A Longitudinal Study of Regional Bracket Equality in the NCAA Men's Basketball Tournament. Saint Michael's College Technical Report: SMC-2013-CS-001. <http://academics.smcvt.edu/jtrono/Papers/BracketStudy.doc>.

West B. (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.