MANAGEMENT SCIENCE
Vol. 32, No. 6, June 1986
Printed in U.S.A.

THE BIAS OF SCHEDULES AND PLAYOFF SYSTEMS IN PROFESSIONAL SPORTS*

HOWARD J. WEISS
School of Business Administration, Temple University, Philadelphia, Pennsylvania 19122

We examine the method used for seeding teams in professional sports leagues after the regular season has ended. In particular, we are interested in which teams are in the playoffs and which of these playoff teams get the home field advantage and/or a first round bye. We show that the regular season schedule is biased against good teams (since teams are separated into divisions) and that in addition to the current schedules and their inherent biases the playoff systems add more bias against good teams (since division winners are typically ranked higher than nondivision winners). The amount of bias in various cases is presented. Furthermore, we present necessary and sufficient conditions for a schedule to be completely unbiased. Using this condition we develop schedules which maximize intradivision rivalries while maintaining no bias.
(SPORTS LEAGUES; SCHEDULING OF TEAMS)

1. Introduction

In this paper we examine the scheduling and playoff systems in professional sports. We show that by dividing leagues into conferences and divisions, and by playing unbalanced schedules the better teams end up with worse records than they would if there were one division and all teams played the same schedule. That this bias exists is not surprising. In this paper we compute the amount of bias under some very reasonable assumptions. The reason for the separation into divisions is to allow for more divisional races. In fact, Whinston and Soni (1982) have shown that pennant races are maximized if no division has four or more teams. However, in any sport there is a penalty to be paid for this divisionalization. For example, in baseball the second best team does not make the playoffs if it is in the same division as the best team in the league. In football, since there are wild card teams the second best team will always make the playoffs. However, this leads to the second part of our analysis which is the seeding of those teams which are in the playoffs.

In 1979 and 1980 the Eagles and Cowboys tied for the best records in the National Football Conference. However in each of those years one of the teams was declared the division winner and the other a wild card team due to complicated tie-breaking rules. The two wild card teams play an extra playoff game and after this extra game they do not play at their home field. This is due to the fact that for playoff purposes the top three teams are the three division winners, even if a wild card team has a better record. The rationale for this is that a division winner may be in a tougher (better) conference than the wild-card team and hence have a worse record. Neither the division winner rule nor its rationale is unique to football. For example, according to this rationale in 1984 the Detroit Tigers and Toronto Blue Jays play in the weak Eastern Division of the American League and get easy wins from their intradivision rivals while Kansas City plays in the more competitive AL West and hence their record suffers from the stiffer competition. The reward for Kansas City is that they are in the playoffs while Toronto is not. Our results (and these examples) show that this rationale is rather weak.

*Accepted by Ambar G. Rao; received May 31, 1984. This paper has been with the author 4 months for 1 revision.

0025-1909/86/3206/0696$01.25

TABLE 1
The Organization of Professional Sports Leagues

Sport	Leagues	Conferences	Divisions
Hockey	NHL (21)	Wales (11)	Patrick (6), Adams (5)
		Campbell (10)	Norris (5), Smythe (5)
Basketball	NBA (23)	Eastern (11)	Atlantic (5), Central (6)
		Western (12)	Midwest (6), Pacific (6)
Football	NFL (28)	National (14)	Eastern (5), Central (5), Western (4)
		American (14)	Eastern (5), Central (4), Western (5)
Baseball	(26)	National (12)	Eastern (6), Western (6)
		American (14)	Eastern (7), Western (7)

Our major result is that we present the amount of bias induced by the schedule and also the additional amount of bias induced by the playoff system (given an already biased schedule). While we do not advocate a change from multi divisions to one division (since it reduces the number of pennant races) we demonstrate that it is clear that the playoff system should be changed so that teams are seeded in the playoff system according to their won/loss records and not whether or not they won the division.

We also show the conditions necessary for won/loss records to have no bias. That is, we show that if the ratio of intraconference games to interconference games does not exceed a certain limit then the won/loss record of a team is an accurate representation of the team's "value" rather than a representation of the schedule that it played. Using this condition we present optimal schedules. By optimal we mean schedules which maximize the number of games played between intradivision rivals and yet maintain the fairness of representing the team "value."

We note that there has been prior work done on ranking (seeding) (for example see Goddard 1983). However, this work is for a post-facto ranking and our goal is to examine seeding rules before play (the season) begins.

In §2 we present our model of the scheduling systems of professional sports. In §3 we present our analysis, definition of bias and results on fair schedules. In §§4 through 7 we analyze baseball, football, basketball and ice hockey.

2. Modelling the Leagues

Essentially, the four major professional sports leagues are divided into two conferences and then each conference is subdivided into two or three divisions. Table 1 contains a list of the leagues, conferences and divisions for baseball, football, basketball and ice hockey along with the number of teams in each league, conference and division. Note that the number of teams in each division does not vary by more than one within any given sport. Furthermore, note that the terminology for baseball is slightly different than for the other sports. That is, the separation into the National and American *Leagues* functions as a separation into *conferences*. Regardless of the reasons for separating leagues into divisions there are two important factors to bear in mind.

Tendency 1. The playoff positions in all sports are awarded according to conferences.

That is, teams in one conference do not compete against teams in a different conference for a playoff berth. Furthermore, in baseball, football and ice hockey, the records of teams in different conferences are not compared for a home field advantage. (Until the 1983 hockey playoffs home ice was accorded in the finals to the team with the better record.) The exception to this tendency occurred in hockey during the years when the first 16 finishers (of 21 teams) were ranked by season record regardless of which conference the team played in. (This is in fact what we advocate!)

The implication of Tendency 1 is that our prime concern is with the relationships of teams within the conference rather than the entire league. The importance of this is that for our computational analysis the problem size is reduced exponentially.

Tendency 2. In each of the four sports, schedules tend to be designed in such a fashion that teams play at least as many games against each team within their division as they play against teams outside their division. (Exact numbers are given in later sections.)

The implication of Tendency 2 is obvious. Teams in strong divisions will play tougher schedules than teams in weak divisions and hence finish the season with lower winning percentages.

Our model is designed to examine the effects due to Tendency 2. Ideally, we would like to model divisions with different numbers of teams as exist in all sports except baseball. However, the analysis will be a combinatorial analysis and assuming equally sized division will reduce the problem size considerably. Therefore we assume that we have a professional sport, that is divided into n divisions with an equal number of m teams in each division. Furthermore, the schedule will have each team play M games against each team within its division and N games against each team outside of the division for a total number of games given by $M(m-1)+N(m)(n-1)$. The specific numbers that will be used for the analysis are given in later sections.

Assume that in any given year each of the $Q = m \cdot n$ teams has a unique innate integer rank between 1 and Q. There are of course $Q!$ ways in which the ranks could be distributed and we assume that each is equally likely. Many of these will yield the same results since each team within any division has the same schedule and we have assumed that divisions are identical in size. Since there are n divisions, each with m teams, the number of arrangements into divisions is given by the number of ways of choosing the m teams for division 1 $\binom{mn}{m}$, multiplied by the number of ways of selecting the m teams for division 2 from the remaining $mn - m$ teams $\binom{mn-m}{m}$ until the last division $\binom{m}{m}$. Furthermore since the division labels do not matter the divisions themselves can be arranged in $n!$ ways. Therefore, the number of permutations that we ultimately need to evaluate is only:

$$R = \binom{mn}{n}\binom{mn-m}{m}\cdots\binom{m}{m}\Big/ n!. \tag{1}$$

Lastly, we need to consider the probability that one team will beat another team. Intuitively, we suspect that the probability that team i will beat team j should be a function of ideally the individual ranks r_i and r_j or, if not, then a function of the difference between the ranks $r_i - r_j$. Define $P(i, j)$ as the probability that team i beats team j. In order to have $P(i, j)$ be both reasonable and general let us presume that $P(i, j)$ is of the form

$$P(i,j) = 0.5 + pf(i,j) \qquad \text{where}$$

$$f(i,j) = -f(j,i), \tag{2}$$

$$f(i,j) \leqslant 0.5 \qquad \text{all } i, j \quad \text{and} \tag{3}$$

$$0 \leqslant p \leqslant 1. \tag{4}$$

Lines (2), (3) and (4) guarantee that $P(i, j)$ is a probability and that $P(i, j)$ and $P(j, i)$ sum to 1. Notice that aside from (2), (3) and (4) we make no other assumptions on

$f(i, j)$. That is, $f(i, j)$ is a general function of the individual teams i and j and $f(i, j)$ may increase or decrease in r_i or r_j or not even be monotone in either variable.

However, while the result that we soon present holds for general $f(i, j)$'s certain function types are more pleasing than others. For example

$$f(i, j) = \begin{cases} 0.5 & \text{if } r_i < r_j, \\ -0.5 & \text{if } r_i > r_j, \end{cases} \tag{5}$$

corresponds to a constant probability. That is, the probability that the higher ranked team wins is given by the fixed constant $0.5(1 + p)$ regardless of how much higher the better team is ranked. A second example is given by

$$f(i, j) = (r_j - r_i)/2(mn - 1). \tag{6}$$

This allows for the probabilities to be functions of the distance between the ranks of the teams as does any function of the form

$$f(i, j) = (r_j - r_i)^z/2(mn - 1)^z \tag{7}$$

for $z = 1, 3, 5, \ldots$. Of course there are an unlimited number of functions which we might choose to serve our purpose. We now prove a result that is both interesting and very useful for our later computations.

THEOREM 1. *Given a specific functional form $f(i, j)$ and the assignment of probabilities according to $P(i, j) = 0.5 + pf(i, j)$ the rankings are independent of the scale factor p if $p > 0$.*

PROOF. For any team i, let $W(i)$ denotes the expected number of wins and let A_i be the set of teams in the same division as team i. Then

$$\begin{aligned} W(i) &= M \sum_{\substack{j \in A_i \\ j \neq i}} (0.5 + pf(i, j)) + N \sum_{j \notin A_i} (0.5 + pf(i, j)) \\ &= 0.5M(m - 1) + pM \sum_{j \in A_i} f(i, j) + 0.5N(Q - m) + pN \sum_{j \notin A_i} f(i, j). \end{aligned}$$

(The first term follows from the fact that there are m teams in the division while the third term follows from the fact that there are $Q - m$ teams outside of the division.)

It follows then that the difference function $D(i, k) = W(i) - W(k)$ is given by

$$D(i, k) = pM \sum_{j \in A_i} f(i, j) + pN \sum_{j \notin A_i} f(i, j) - pM \sum_{j \in A_k} f(k, j) - pN \sum_{j \notin A_k} f(k, j). \tag{8}$$

As the values for $f(i, j)$ are fixed it follows that p can be factored out and that the sign of $D(i, k)$ changes only when $p < 0$. Hence the result is proved.

The major implication of this result is that for the computational work that follows we can always assume that $p = 1$ without any loss of generality.

In order to keep the exposition simple assume that the higher ranked team always wins. That is, $f(i, j)$ is given by (5) and $p = 1$. It follows that the expected number of wins for each team is given by

$$w(r, d) = M(m - d) + N(mn - m + d - r), \tag{9}$$

where $r \in \{1, 2, \ldots Q\}$ is the rank of the team within the conference and $d \in \{1, 2, \ldots, m\}$ is the rank of the team within the division.

Now, for any particular seeding system we can compute the number of wins and the playoff seedings according to that seeding system. Foremost among the seeding systems is the one we advocate (which has been used in hockey in the past)—seed teams in the playoffs according to the number of wins. Another popular (but very biased) system is giving the top seeds to the division winners according to the number of wins and then seeding the remaining teams by number of wins (as used in football, basketball and baseball).

Consider for example a league which consists of three divisions, four teams per division and $M = 4$ and $N = 1$. Suppose that the innate rankings are given as follows:

	Division											
	Eastern				Midwestern				Western			
Team	D	P	W	S	C	M	G	T	A	L	N	SF
Rank	1	2	3	4	5	6	7	8	9	10	11	12

Then according to the formula for wins (9) the following table can be computed:

Team	D	P	S	W	C	M	G	T	A	L	N	SF
Number of Wins	20	16	12	8	16	12	8	4	12	8	4	0
Seeding by Wins	1	2(t)	4(t)	7(t)	2(t)	4(t)	7(t)	10(t)	4(t)	7(t)	10(t)	12
Seeding by Division Winner System	1	4	5(t)	7(t)	2	5(t)	7(t)	10(t)	3	7(t)	10(t)	12

In this example the second best team in the conference ends up with a tie for the second best record rather than the unique second best record because it is in a stronger division. Furthermore, according to the division winner system it is placed in the fourth playoff position because it did not win the division. Of course, when one team is penalized another team is rewarded but our computational results will show that the penalties and rewards are distributed unfairly.

3. The Season Schedule

3.1. *Evaluation Criteria for Divisional Play*

We will enumerate the results of all of the possible permutations of ranks for different values of m, n, M and N to represent the four sports. For each set of enumerations we are interested in determining the relationship between rank and seed. Define

$a(i, j)$ = the number of times (out of the R permutations given by (1)) that the team that has innate rank j is seeded in position i for the playoffs under a particular system of seeding.

In an ideal system the team that has innate rank j should end up as the jth seed after the season. Therefore, for any cell i, j define bias (of the cell) as the difference $(j - i)$ between the rank (j) and the seed (i). Then for any seeding system define in the usual statistical sense the following biases and consistencies:

B_j = Bias of rank $j = \sum_{i=1}^{mn}(j-i)p_{ij}$,
b_i = Bias of position $i = \sum_{j=1}^{mn}(j-i)p_{ij}$,
C_j = Consistency of rank $j = \sum_{i=1}^{mn}(j-i)^2p_{ij}$,
c_i = consistency of position $i = \sum_{j=1}^{mn}(j-i)^2p_{ij}$,

where $p_{ij} = a_{ij}/R$, and R is the total number of permutations as given by (1). Also define total consistency as $S = \sum_{j=1}^{mn} C_j = \sum_{i=1}^{mn} c_i$. (We do not define total bias since this is equal to zero by nature of the assumptions.)

We now present the necessary and sufficient condition for the record of each team to be identical with its rank, under different assumptions about $P(i,j)$. Define such a schedule as completely consistent. That is a schedule is completely consistent if its total consistency S is equal to zero. This implies that under any permutation of ranks the rank of each team is equal to its seed (rank of its won/loss record). We have the following.

THEOREM 2. a. *Under the assumption of constant probabilities (that is, when $f(i,j)$ is given by* (5)) *a schedule is completely consistent if and only if $M \leqslant Nm(m-1)$.*

b. *Under the assumption of linear probabilities (that is, when $f(i,j)$ is given by* (6)) *a schedule is completely consistent if and only if $M \leqslant N(m^2n - m^2 - m + 2)/(m^2n - m^2 - m + 2 - mn)$.*

PROOF. Consider two teams i and k and without loss of generality assume that team i has a better rank (i.e. $r_i < r_k$). For either part of the theorem we wish to consider the difference function $D(i,k) = W(i) - W(k)$. Now, under the assumption that $f(\cdot,\cdot)$ is constant the worst case (i.e. the one most likely to have team k accumulating more wins than team i) is when team k is only marginally worse than team i (i.e. $r_k = r_i + 1$) and when team i is last in its division while team k is first in its division. Therefore the worst of the permutations appears as follows:

Division A	Division B
team 1	team $k = i + 1 = m + 1$
team 2	team $(n-1)m + 2$
⋮	⋮
team $i = m$	team $(n-1)m + m$

(By team j we mean the team with rank j.) The missing teams, team $m+1$ through team $(n-1)m+1$, are scattered over the remaining $n-2$ divisions in any combination. Notice that team i is in the best division and team k is in the worst division.

In this case $W(i) = Nm(m-1)$ while $W(k) = M(m-1) + m(n-2)N$. $D(i,k) = Nm - M(m-1)$ which is greater than zero if and only if $M < Nm/(m-1)$ and the first part of the theorem is complete.

b. In this case, since the probability of winning increases with the differences in ranks the worst case is when team 1 is in the best division and team 2 is in the worst. That is, the divisions are given by

Division A	Division B
team 1	team 2
team 3	team mn
⋮	⋮
team $m+1$	team $mn - m + 1$

The difference function $D(i,k)$ is given by (8) and from Theorem 1 we can exclude p. Furthermore since we only care about the sign of $D(i,k)$, we can exclude the term $2(n-1)$ which is in the denominator of $f(i,j)$. Hence

$$D(1,2) = \sum_{j=2}^{m} j + N + N \sum_{m+1}^{mn-1} j - \left(M \sum_{j=m(n-1)}^{mn-2} j - N + N \sum_{j=1}^{m(n-1)-1} j \right)$$

which after much algebra becomes

$$M(m^2 - m^2n + m - 2 + mn) + N(m^2n - m^2 - m + 2).$$

Since team 1 is better than team 2 we require that $D(1,2) > 0$ and b is proved.

3.2. *Optimal Schedules*

Theorem 2 raises the following problem of maximizing games between classic rivalries while still yielding fair results. That is, given m, n and a total number of games T find the values of M and N such that M (the number of intradivision games) is maximized and the condition of Theorem 2 is satisfied. Consider the constant probability case. The problem can be expressed as:

$$\begin{aligned} &\max \quad M \\ &\text{s.t.} \quad M < Nm/(m-1) \qquad \text{(Theorem 2)}, \\ &\qquad M(m-1) + N(n-1)m = T \qquad \text{(total number of games)}. \end{aligned}$$

Now M and N do not have to be integers. For example, if divisions have 6 teams then $M = 7\frac{4}{5}$ is interpreted as play four teams 8 times and 1 team 7 times. Therefore M must be an integer multiple of $1/(m-1)$ while N must be an integer multiple of $1/m(n-1)$. Define new variables k and c as given by the total number of intradivision and interdivision games respectively. That is, $k = M(m-1)$ and $c = m(n-1)N$.

After substitution the problem becomes

$$\begin{aligned} &\max \quad k \\ &\text{s.t.} \quad k < c/(n-1), \\ &\qquad k + c = T, \\ &\qquad k, c \text{ nonnegative integers.} \end{aligned}$$

Since c and k are integers the first constraint can be rewritten as an equality by subtracting 1 from the right-hand side yielding $k \leqslant c/(n-1) - 1$. The problem does not necessarily have a solution. However if $T + 1$ is a multiple of n then the solution is obtained as

$$k = (T+1)/n - 1, \quad c = (T+1)(n-1)/n \quad \text{or}$$

$$M = (T+1-n)/n(m-1), \quad N = (T+1)/nm$$

In Table 2 we present the optimal schedules for the four sports under assumptions that resemble the structure of each. Consider baseball for example. Each conference (league) has 2 divisions with 6 or 7 teams. The total number of games in baseball is 162 and we set $T = 161$ since we need $T + 1$ to be a multiple of n. In the National League, the schedule that maximizes intradivision competition and remains fair would have each team play 16 games against each intradivision rival, 13 games against 3 of the rivals in the other division and 14 games against the remaining three rivals outside of the division. (The current values are $M = 18$ and $N = 12$.) The results for the remaining sports can be found in Table 2.

TABLE 2

Optimal Schedules under Equal Probability Assumption
(Maximize rivalries subject to fairness)

Sport	n	m	T	M	N
Baseball (NL)	2	6	161	16	$13\frac{3}{6}$
Baseball (AL)	2	7	161	$13\frac{2}{6}$	$11\frac{4}{7}$
Football	3	5	14	1	1
Basketball	2	6	55	$5\frac{2}{5}$	$4\frac{4}{6}$
Ice Hockey	2	5	55	$6\frac{3}{4}$	$5\frac{3}{5}$

In Table 3 we present a similar analysis except under the assumption of linear probabilities. We have used the same parameters (n, m and T) as for the constant probability assumption but we have no guarantee that the answers are integer. The results are derived by solving the simultaneous equations

$$T = M(m-1) + Nm(n-1) \qquad \text{(total games)},$$

$$M = N(m^2n - m^2 - m + 2)/(m^2n - m^2 - m + 2 - mn) \qquad \text{(Theorem 2)}.$$

Notice from Tables 2 and 3 that any schedule that satisfies the linear probability assumption satisfies the equal probability assumption. That is, in terms of fairness, scheduling requirements are stricter under the assumption of equal probabilities rather than the assumption of linear probabilities.

We complete our discussion of fairness with one final result.

COROLLARY 3. *Under the equal probability assumption the jth best team according to innate rank will have the jth best record under any permutation of ranks if and only if* $M \leqslant Nj(j-1)$ *probability assumption.*

PROOF. The proof follows by substitution of j for m in the proof of Theorem 2, part a.

At this point we are ready to present our computational results. For each sport the results that we present are for the equal probability assumption. The reason for this is that as has been shown the equal probability assumption is stricter than the linear probability assumption. It follows that assuming equal probabilities will generate worst case results for the different sports. We make one final remark. While intuitively we feel that $P(i, j)$ should increase as $r_j - r_i$ increases we have not been able to disprove the constant probability assumption (using a χ^2 goodness of fit test) for data from past baseball seasons.

4. Baseball

Baseball is the easiest sport to analyze for several reasons. It is the only sport in which teams from one conference do not play teams from another conference.

TABLE 3

Optimal Schedules under Linear Probability Assumption

Sport	n	m	T	M	N
Baseball (NL)	2	6	161	18.40	11.50
Baseball (AL)	2	7	161	14.95	10.19
Football	3	5	14	1.30	0.88
Basketball	2	6	55	6.29	3.93
Ice Hockey	2	5	55	8.18	4.46

Furthermore, in the National League all teams play the same schedule (18 games against each team within the division and 12 games against each team outside the division). In the American League each team plays 15 games against teams within the division, 10 games against five of the seven teams outside the division and 11 games against two of the seven teams outside of the division. Therefore, we use the following parameters for our analysis:

National League: $m = 1$, $n = 6$, $M = 18$, $N = 12$,

American League: $m = 2$, $n = 7$, $M = 15$, $N = 10\frac{2}{7}$.

National League

Table 4 contains the values of a_{ij} for the National League. (In the cases of ties the positions (i) were spread evenly over the number of ties. For example, in the example of §2, $\frac{1}{3}$ was added to each of $a(4,3)$, $a(5,3)$ and $a(6,3)$ since there was a three-way tie for 4th and 5th and 6th places according to number of wins.) Compared with numbers that will appear later the a_{ij}'s are very reasonable. Seeding the teams by number of wins is never more than one position away from the innate rank of the team except in one (tied) case of the 462 permutations (see a_{57}). Furthermore, the top two innately ranked teams always end up in the proper rank by records which we know from Corollary 3.

The current system in baseball (and all other sports) does not rank teams by season record. As baseball operates now the two division winners make the playoffs. In Table 5 we present the results (a_{ij}) using the division winner rule. This leads to obvious problems. For example whenever the two best teams are in the same division one of them does not make the playoffs. This happens $\frac{5}{11}$ of the time. In fact, in a league with $m \cdot n$ teams, m in each division, the probability that the second best team is in the same division as the best team is clearly $(m - 1)/(mn - 1)$. Thus, the baseball division winner rule clearly penalizes good teams and helps mediocre teams as it enables the third through seventh ranked teams to enter the playoffs with probabilities of 126/462, 56/462, 21/462, 6/462 and 1/462 respectively.

During the strike year in baseball (1981) there were four teams from each league in the playoffs. Since that time there has been discussion about expanding the playoffs to four teams permanently. The effects of this expansion can be seen from Table 5. Using the division winner rule the top three teams will always be ranked one through four and make the playoffs. The fourth best team would miss the playoffs only the 28/462

TABLE 4

National League: Seeding by Number of Wins

	Innate Rank											
	1	2	3	4	5	6	7	8	9	10	11	12
Seed												
1	462	0	0	0	0	0	0	0	0	0	0	0
2	0	462	0	0	0	0	0	0	0	0	0	0
3	0	0	434	28	0	0	0	0	0	0	0	0
4	0	0	28	413	21	0	0	0	0	0	0	0
5	0	0	0	21	404.5	36	0.5	0	0	0	0	0
6	0	0	0	0	36.5	400	25.5	0	0	0	0	0
7	0	0	0	0	0	25.5	400	36.5	0	0	0	0
8	0	0	0	0	0	0.5	36	404.5	21	0	0	0
9	0	0	0	0	0	0	0	21	413	28	0	0
10	0	0	0	0	0	0	0	0	28	434	0	0
11	0	0	0	0	0	0	0	0	0	0	462	0
12	0	0	0	0	0	0	0	0	0	0	0	462

TABLE 5

National League: Seeding Division Winners First

	Innate Rank 1	2	3	4	5	6	7	8	9	10	11	12
Seed												
1	462	0	0	0	0	0	0	0	0	0	0	0
2	0	252	126	56	21	6	1	0	0	0	0	0
3	0	210	252	0	0	0	0	0	0	0	0	0
4	0	0	84	378	0	0	0	0	0	0	0	0
5	0	0	0	28	404	30	0	0	0	0	0	0
6	0	0	0	0	37	400	25	0	0	0	0	0
7	0	0	0	0	0	25.5	400	36.5	0	0	0	0
8	0	0	0	0	0	0.5	36	404.5	21	0	0	0
9	0	0	0	0	0	0	0	21	413	28	0	0
10	0	0	0	0	0	0	0	0	28	434	0	0
11	0	0	0	0	0	0	0	0	0	0	462	0
12	0	0	0	0	0	0	0	0	0	0	0	462

times that it was in the same division as the best three teams. Hence, we must advocate (on the basis of fairness) that baseball expand its playoff system to include each of the two division winners and the two teams with the best records of the remaining teams. While this four-team system is fairer than what currently exists we note that if four teams are to be chosen then it is still preferable to choose them according to won-loss record as this yields only 21 times that the fourth best team does not make the playoffs (see Table 4).

As we indicated before bias and consistency are relevant measures for ranking procedures. In Table 6 we present the bias by team and the bias by playoff position for each of the two ranking systems examined. The interpretation of the team bias (B_j) is that it is the average deviation of the playoff seed from what the innate rank is. For example, according to won/loss record, on average team 3 is seeded 0.061 below (sign is negative) the rank of 3 while according to division winners first team 3's seed increases by 0.09 over its rank of 3.

Lastly, we report that the consistency for the two methods are 0.857 for seeding by record and 2.753 for seeding division winners first. As the bias totals zero, clearly consistency is a highly appropriate measure of the seeding systems and the indication

TABLE 6

National League BIAS

	Seed by record		Seed by Division Winners First	
Team (j) or Seed (i)	B_j = Bias of team j	b_j = Bias of seed i	B_j = Bias of team j	b_j = Bias of seed i
1	0	0	0	0
2	0	0	− 0.455	0.714
3	− 0.061	0.061	0.090	− 0.455
4	0.015	− 0.015	0.182	− 0.182
5	− 0.034	0.035	0.056	0.004
6	0.021	− 0.024	0.060	− 0.026
7	− 0.021	0.024	− 0.013	− 0.024
8	0.034	− 0.035	0.034	− 0.035
9	− 0.015	0.015	− 0.015	0.015
10	0.061	− 0.061	0.061	− 0.061
11	0	0	0	0
12	0	0	0	0

is that seeding by records is more than three times as consistent as ranking division winners first.

American League

The results for the American League are summarized in Tables 7 and 8. They are very similar to the results on the National League however there are a few observations to make. Corollary 3 notes the importance of the ratio of M to N. The smaller M/N is the less are the effects of dividing leagues into divisions. Since the American League has a ratio of $15/10\frac{2}{7}$ compared with the ratio of $3/2$ for the National League the ranking by record is more consistent. Furthermore Corollary 3 tells us that the top three teams always end up with the top three won loss records. This is verified in Table 7. In fact the actual consistencies are 0.570 for ranking by record and 2.761 for ranking division winners first.

5. Football

The 28 teams in the NFL are divided into two conferences and each of these conferences is divided into three divisions with five, five and four teams. The schedule consists of 16 games for each team and is unbalanced in the sense that each team plays each team within its own division exactly two times for a total of 8 (or 6) games and then plays 8 (or 10) teams outside of its division exactly one time. We would like to take out the variation that occurs according to whether a team is in a four- or five-team division and also whether it was fifth or not in the division. Therefore for our purposes we assume that each conference consists of three four-team divisions and that all games are played within the conference.

Under the current 14-team conference system teams play other teams within their division two times and play two teams from each of the other two divisions in the same conference. In any given year which of the two teams is played is random. (It is determined by the previous year's results. Therefore, while the schedule for 1986 is known today it is impossible to predict which teams will play in 1987.) Since we do not know which of the two teams will be on the schedule, then from an expected value viewpoint that is the same as playing each team exactly one halftime. Thus in our football model, $n = 3$, $m = 4$, $M = 2$, $N = 1/2$. (The number of teams in each division is either 4 or 5 and we have chosen to set $m = 4$ rather than 5 simply because it makes the tables easier to read.)

Table 9 contains the results of seeding by record. Notice that the bias by team is a monotone function while the bias by position is very ill-behaved. The interpretation is as follows: the team that ends up with the third best record is on average 3453/5755 above rank 3. Notice that on average, position 4 is the one out of line with respect to having a better team than the teams record would indicate. That is, going to the original data (Table 9) approximately one half of the time (805 + 2020)/5775 the team with the fourth best record is actually the second or third best team.

Evaluation of the Playoff System

For the same set of 5755 orderings we have positioned the teams as the NFL does. That is, the division winners are placed in the top three positions according to their records and then the remaining nine teams are placed in positions four through twelve according to their records. The results including bias and consistency appear in Table 10. One immediate observation is that this current system is less consistent then simply ranking the teams by their won/loss record. A second observation is that the (negative) team bias has gotten worse (in the sense that they are more negative) for the second, third and fourth best teams and gotten better or stayed the same for the

TABLE 7

American League: Seeding by Record
Consistency = 0.570

	Innate Rank														Seed Bias
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	b_i
Seed															
1	1716	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1716	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1716	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	1632	84	0	0	0	0	0	0	0	0	0	0.049
5	0	0	0	84	1604	28	0	0	0	0	0	0	0	0	− 0.033
6	0	0	0	0	28	1575	112	1	0	0	0	0	0	0	0.050
7	0	0	0	0	0	113	1567	36	0	0	0	0	0	0	− 0.045
8	0	0	0	0	0	0	36	1567	113	1632	0	0	0	0	0.045
9	0	0	0	0	0	0	1	112	1577	28	0	0	0	0	− 0.050
10	0	0	0	0	0	0	0	0	28	1604	84	0	0	0	0.033
11	0	0	0	0	0	0	0	0	0	84	1632	0	0	0	− 0.049
12	0	0	0	0	0	0	0	0	0	0	0	1716	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	1716	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1716	0
B_j (Team Bias)	0	0	0	− 0.049	0.033	− 0.050	0.043	−0.043	0.050	− 0.033	0.049	0	0	0	

TABLE 8

American League: Seeding Division Winners First

Consistency = 2.761

	Innate Rank														Seed Bias
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	b_i
Playoff Seed															
1	1716	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	924	462	210	84	28	7	1	0	0	0	0	0	0	0.750
3	0	702	924	0	0	0	0	0	0	0	0	0	0	0	− 0.462
4	0	0	330	1386	0	0	0	0	0	0	0	0	0	0	− 0.192
5	0	0	0	120	1596	0	0	0	0	0	0	0	0	0	− 0.070
6	0	0	0	0	36	1575	105	0	0	0	0	0	0	0	0.040
7	0	0	0	0	0	113	1567	36	0	0	0	0	0	0	− 0.045
8	0	0	0	0	0	0	36	1567	113	0	0	0	0	0	0.045
9	0	0	0	0	0	0	1	112	1575	28	0	0	0	0	− 0.050
10	0	0	0	0	0	0	0	0	28	1604	84	0	0	0	0.033
11	0	0	0	0	0	0	0	0	0	84	1632	0	0	0	− 0.049
12	0	0	0	0	0	0	0	0	0	0	0	1716	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	1716	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1716	0
B_j (Team Bias)	0	− 0.462	0.077	0.175	0.126	− 0.001	0.059	−0.041	0.048	− 0.033	0.049	0	0	0	

TABLE 9

National Football League: Seeding by Record

Consistency = 12.996

	Innate Rank												Seed Bias
	1	2	3	4	5	6	7	8	9	10	11	12	b_i
Playoff Seed													
1	5775	0	0	0	0	0	0	0	0	0	0	0	0
2	0	4217.50	1260	280	17.50	0	0	0	0	0	0	0	0.324
3	0	752.50	2180	1680	962.50	200	0	0	0	0	0	0	0.598
4	0	805	2020.67	1860	281	502.17	260	44	2.17	0	0	0	− 0.239
5	0	0	136.17	1140	2904	1198.17	270	96	30.67	0	0	0	0.128
6	0	0	178.17	645	678	1649.67	1695	792	137.17	0	0	0	0.206
7	0	0	0	137.17	792	1695	1649.67	678	645	178.17	0	0	− 0.206
8	0	0	0	30.67	96	270	1198.17	2904	1140	136.17	0	0	− 0.128
9	0	0	0	2.17	44	260	502.17	281	1860	2020.67	805	0	0.239
10	0	0	0	0	0	0	200	962.50	1680	2180	752.50	0	− 0.598
11	0	0	0	0	0	0	0	17.50	280	1260	4217.50	0	− 0.324
12	0	0	0	0	0	0	0	0	0	0	0	5775	0
B_j (Team Bias)	0	− 0.409	− 0.271	− 0.127	− 0.081	− 0.037	0.037	0.81	0.127	0.271	0.409	0	

remaining teams. This observation is critical as it means the following: *Given the current divisional set-up and schedule none of the better teams benefits from the playoff format when compared to a format based solely on won/loss record.*

Just as interesting though is the observation that regardless of which playoff system is chosen the same teams reach the playoffs with but an exception of 2/5575. (If the playoffs contained the top six teams rather than five teams then the same teams would always be in the playoffs whether chosen by record or division winner.)

6. Basketball

The NBA has 23 teams which are divided into two conferences of 11 and 12 teams. As the number of teams in each division varies the schedule of each team varies slightly. In order to have a schedule to analyze we use the schedule of the 5-team Atlantic Division. Each team plays 6 games against each of its Atlantic Division rivals, 5 games against four of the six other conference rivals and 6 games against the remaining two conference rivals: They also play two games against each of the teams in the Western Conference. Playoff seeding in basketball is on a conference basis which means that our concern is with respect to the conference rather than the league. Therefore, for our basketball analysis we let $n = 2$, $m = 6$, $M = 6$, $N = 5\frac{4}{6}$.

It is not necessary to present the table of seeding by won-loss record because basketball satisfies Theorem 2. That is, $M = 6 \leqslant Nm/(m-1) = (34/6)(6)/(5) = 6.8$; therefore, seeding by record is completely unbiased.

For the past several years basketball has seeded teams in the playoffs according to the division winner rule. Prior to 1984 the division winners received byes and for each conference the next four teams were chosen according to won/loss record. Subsequently, home court advantage was given according to won/loss record. Thus, in basketball, a division winner could have received the advantage of a bye but then home court advantage was done properly. Table 11 contains the results for basketball's division winner advantage.

7. Hockey

The regular season schedule and the playoff system in hockey has varied a great deal since the expansion of the league from its original six teams. During the 1979/80 season each team played the same schedule (4 games against every other team) and the top 16 teams are seeded for the playoffs according to their won/loss record. No system could be fairer than this. In recent years the playoff positions were awarded according to division position (rather than conference position). Hence hockey as the system exists now is an aberration from our model. That is, rather than having conference seeds 1 through 8 there are two groups of divisional seeds 1 through 4. For this reason we have not performed our combinatorial analysis. We do note that since the schedule is unbalanced the won/loss records cannot be compared across divisions but that within the division the seeding is unbiased.

8. Summary

The analysis we have performed has demonstrated several facts. Some of these facts may have been surmised before the analysis but are documented in this paper.

1. Divisional play is disadvantageous for the better teams in that their won/lost records are worse than if every team played an identical schedule. The exception to this rule is basketball in which case teams within the conference essentially play each other an equal number of times.

TABLE 10

National Football League: Seeding by Division Winners First

Consistency = 15.043

	Innate Rank 1	2	3	4	5	6	7	8	9	10	11	12	Seed Bias b_j
Playoff Seed													
1	5775	0	0	0	0	0	0	0	0	0	0	0	0
2	0	4200	1260	280	35	0	0	0	0	0	0	0	0.333
3	0	0	1680	1680	1190	700	350	140	35	0	0	0	1.467
4	0	1575	2520	1680	0	0	0	0	0	0	0	0	− 0.982
5	0	0	135	1320	2940	1200	180	0	0	0	0	0	− 0.005
6	0	0	180	645	678	1650	1695	792	135	0	0	0	0.204
7	0	0	0	137.17	792	1695	1649.67	678	645	178.17	0	0	− 0.206
8	0	0	0	30.67	96	270	1198.17	2904	1140	136.17	0	0	− 0.128
9	0	0	0	2.17	44	260	502.17	281	1860	2020.67	805	0	0.239
10	0	0	0	0	0	0	200	962.50	1680	2180	752.50	0	− 0.598
11	0	0	0	0	0	0	0	17.50	280	1260	4217.50	0	− 0.324
12	0	0	0	0	0	0	0	0	0	0	0	5775	0
B_j (Team Bias)	0	−0.545	−0.358	− 0.158	− 0.042	0.049	0.113	0.122	0.139	0.271	0.409	0	

TABLE 11

National Basketball Association: Seeding by Division Winners First

Consistency = 2.14

	Innate Rank												Seed Bias
	1	2	3	4	5	6	7	8	9	10	11	12	b_i
Playoff Seed													
1	462	0	0	0	0	0	0	0	0	0	0	0	0
2	0	252	126	56	21	0	1	0	0	0	0	0	0.714
3	0	210	252	0	0	0	0	0	0	0	0	0	− 0.455
4	0	0	84	378	0	0	0	0	0	0	0	0	− 0.182
5	0	0	0	28	434	0	0	0	0	0	0	0	− 0.061
6	0	0	0	0	7	455	0	0	0	0	0	0	− 0.015
7	0	0	0	0	0	1	461	0	0	0	0	0	− 0.002
8	0	0	0	0	0	0	0	462	0	0	0	0	0
9	0	0	0	0	0	0	0	0	462	0	0	0	0
10	0	0	0	0	0	0	0	0	0	462	0	0	0
11	0	0	0	0	0	0	0	0	0	0	462	0	0
12	0	0	0	0	0	0	0	0	0	0	0	462	0
B_j (Team Bias)	0	− 0.455	0.091	0.182	0.121	0.050	0.011	0	0	0	0	0	

2. If schedules satisfied Theorem 2 then divisional play would give a perfect indication of the teams' actual strengths rather than scheduling advantages/disadvantages.

3. The current playoff ranking system of having the top ranked teams be the divisional winners is disadvantageous to the better teams.

In summary, the professional sports leagues would operate in a fairer fashion, yet still have pennant races, if the schedule satisfied Theorem 2 and playoff seedings were according to won/loss record.

References

GODDARD, S. T., "Ranking in Tournaments and Group Decision Making," *Management Sci.*, 29, 12 (December 1983), 1384–1392.

WINSTON, W. AND A. SONI, "Does Divisional Play Lead to More Pennant Races?," *Management Sci.*, 28, 12 (December 1982), 1432–1440.