Powerwise vs. the RPI

Comparing Powerwise with the current method for at-large selection to the MD1 lacrosse championship tournament.

Authors: Lawrence Feldman and Matthew Bomparola

Abstract

The purpose of this paper is to compare Powerwise and its Power Rating algorithm against historical ranking lists generated by the lacrosse selection committee with statistical support from the Ratings Percentage Index (RPI). It will show that despite, and as a virtue of, its simplicity, Powerwise produces common-sense results that compete with and improve on the existing committee + RPI approach to at-large selection.

The primary tool used to complete this analysis was inspired by Stocks-Smith and consists of plotting game scores vs. opponent strength for any discrepancies between Powerwise and historical at-large picks. Other methods include Barrow et al.'s rank correlation analysis, Paul and Wilson's probit bias analysis, and a transitivity analysis inspired by S.I. Gass.

Keywords: At-large selection; Men's Division I Lacrosse; Pairwise Comparisons; Power Ratings; Goal differences and margin of victory, on-field results; Sagarin, Massey, Colley.

Introduction

The lacrosse coaches' committee set out to develop a rating system uniquely suited to selecting teams for the NCAA Men's Division I lacrosse championship tournament. The committee examined over a dozen potential methods, testing each against 10 years of historical data and years of committee-member experience before settling on Powerwise as the top candidate.

The committee dreamed of a method that would replace the existing system for rating teams, which relies heavily on the biased and mathematically-flawed Ratings Percentage Index (RPI) and closed-door committees that—despite producing reasonable results—feel like a "byzantine black box," leading to a "large amount of speculation, second guessing, and debate each year about [committee] decisions." (Coleman et al., 2010; Colley, 2002)

Powerwise is an attractive potential replacement thanks to few key factors, including:

I. Emphasis on head-to-head and common opponents' records.
II. Utilization of information-rich margins of victory.
III. Minimal chance for human intervention.
IV. Transparency and simplicity.
V. Common-sense results.

Powerwise's hierarchical approach seems to solve the challenges of designing a rating system by aligning rigorous statistics with the common-sense intuition that on-field results should outweigh mathematical metrics whenever possible.[1]

But does it produce reasonable results?

Literature Review

Sports ranking is a difficult endeavor because, as cites in the whitepaper introducing the Colley Matrix, "there is no ranking that is an absolute truth, against which to check [your hypothesis]." (Colley, 2002) Thankfully, the extensive literature on comparing sports ranking systems answered many questions that surfaced during the development of this analysis, including:

I. To what extent should on-field results outweigh statistical methods?
II. To what extent do margins of victory improve ranking accuracy?
III. How might one test for bias in favor of certain teams or conferences?
IV. Which mathematical characteristics signal a strong algorithm?
V. How might one compare two different at-large picks?

In general, Powerwise's use of margins of victory, its preference for on-field results, and its emphasis on simplicity and transparency are well justified by the existing literature.

[1] Powerwise is described in full in Appendix A.

Analysis performed by Barrow et al. and Annis and Craig, for instance, finds that methods designed to use score-differential data tend to be both more predictive and more likely to converge on an interpretable, statistically-significant ranking solution than those that use only win-loss records. Barrow's analysis of real data shows "(not surprisingly) that there is some information conveyed in scores that is not captured in wins and losses." (Annis and Craig, 2005; Barrow et al., 2013)

Information is immensely valuable in a sports ranking context where many teams play relatively few games. Methods that ignore interpretable data risk producing incorrect or systematically biased results. Confirming this suspicion, Paul and Wilson's probit analysis builds on an earlier analysis by Coleman et al., identifying the existence of systematic bias in favor of major NCAA basketball conferences and attributing it to the RPI's exclusion of margin-of-victory data.(Coleman et al., 2010; Paul and Wilson, 2015)

Suspicion of bias in favor of certain teams or conferences is a mainstay in NCAA sports. Colley and Coleman et al. point out the obvious: the secrecy and opaqueness of closed-door selection committees and "byzantine black-box" analytics give coaches, players, and fans "little reason to reconsider their doubts." In light of this, many (including the lacrosse coaches' committee) believe that, when it comes to stats, the simpler the better. (Coleman et al., 2010; Colley, 2002)

The lacrosse selection committee is not alone in expressing a preference for on-field results. In 2018, Mountain West football, for instance, moved "to place a priority on head-to-head competition" and NCAA Hockey already uses a "sister" method to Powerwise that inspired its use of hierarchical pairwise comparisons that place emphasis on head-to-head and common opponents' records.(*San Diego Union-Tribune*, 2018; *USCHO.com | FAQ | Frequently Asked Questions*, n.d.)

Other than those mentioned, this analysis benefitted immensely from analyses and whitepapers published by Ray Stefani, Peter Juma Ochieng András London and Miklós Krész, Justin Stocks-Smith, and Kenneth Massey (among others). The ranking methods described and tested therein inspired the creation of Powerwise and this analysis.(Massey, 1997; Ochieng et al., 2022; Stefani, 2011; Stocks-Smith, 2021)

Powerwise and the RPI

To determine a Powerwise ranking list, each team is compared against every other team in a "hypothetical" pairwise matchup. The winner of each matchup is awarded a "Powerwise point." Determining the winner of each matchup is done by a three-step hierarchical procedure that emphasizes real-world, on-field performance.

Winning a matchup is usually (>60% of the time) a matter of having a superior head-to-head record, or, if that record is tied or nonexistent, a better record against common opponents. When neither on-field metric is decisive, Powerwise defaults to a tried-and-true goal-differential-based statistic inspired by Sagarin's "Predictor" called the Power Rating. Once all Powerwise points have been awarded, teams are ranked based on the number of points they received.

For a more detailed explanation and information about the RPI, consult Appendix A and B.

Data and Methods

To aid this analysis, ten years of historical games and rating data were assembled from a variety of sources, including, most notably, Feldman's repository hosted at laxmath.com.

i. Methods

At the heart of this analysis are a visualization of all discrepancies between historic at-large selections and hypothetical Powerwise-based picks.

In their whitepaper on the CBR, Stocks-Smith identified six discrepancies in at-large picks (of 36 total) between the extant committee-based basketball selection method and their new method, concluding that the CBR produced superior at-large picks. One point of comparison was a plot of all games' score differentials (y-axis) against the opponents' strength (x axis), effectively visualizing a team's performance against their strength of schedule. (Stocks-Smith, 2021)

Borrowing this visualization (and expanding on it with linear regression) informs our analysis of the discrepancies between Powerwise and the lacrosse selection committee over the past ten years.

Other supporting analysis includes rank correlation analysis by way of Kendall's tau; a comparison of real-world transitivity in games results as predicted by both Powerwise and the (correlated-with-official) RPI ranking list, as inspired by S. I. Gass; and a conference bias probit model inspired by Paul and Wilson testing the effect of team strength as captured by Powerwise or the RPI (independent variable) on whether a given team really received an at-large bid (dependent variable), with dummy variables for conference size identifying any residual effect caused by conference affiliation. (Barrow et al., 2013; Paul and Wilson, 2015; S. I. Gass, 1998)

ii. Limitations

This paper attempts to test Powerwise in a variety of ways, but time is an ever-powerful constraint. Comprehensive analysis of Powerwise's stability under perturbation, its predictive power, its time-to-convergence, and its constituent Power Rating (PR) statistic are left to future research.

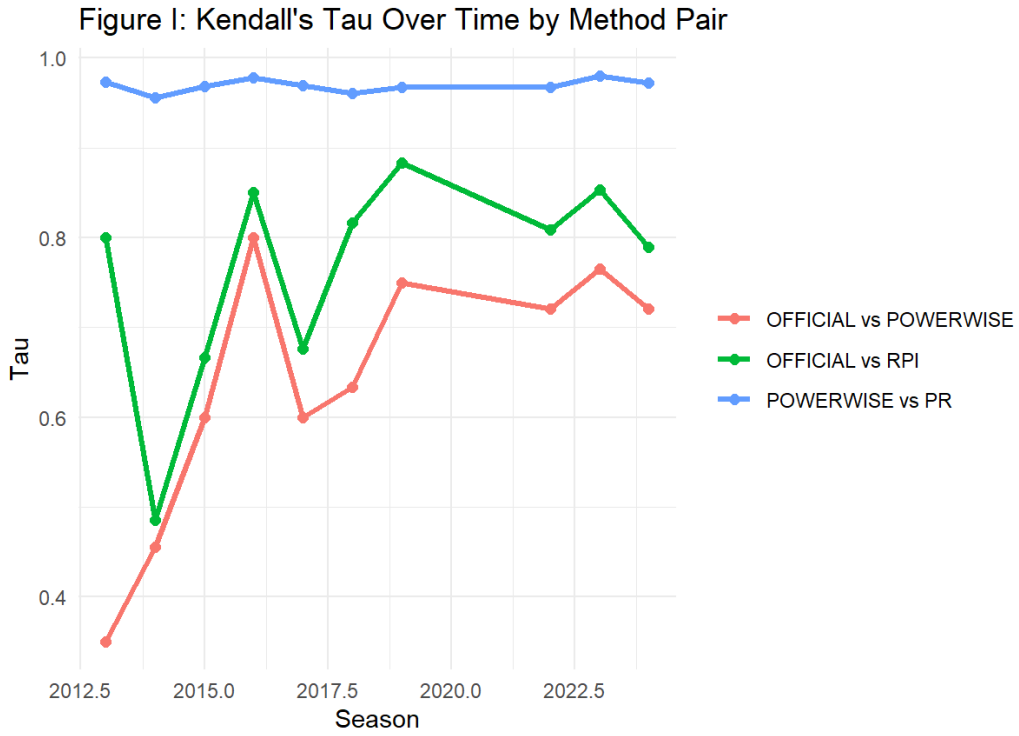
Testing predictive strength is particularly difficult and somewhat out-of-scope of this analysis. That being said, Barrow et al. and others' comprehensive analysis of the predictive strength of other popular methods (including a least-squares method similar to the Power Rating) fuel hope that Powerwise stands a chance against the best.

Regardless, predictive accuracy tests for the *strongest* or *best* teams at a given point in a season. Powerwise takes another approach; rather than attempting to model "true" team strength, it estimates *deservedness*, a quality better aligned with the values of fairness and on-field merit emphasized by the coaches' committee.

Results

i. Kendall's tau

Figure 1 shows that official rankings and RPI ratings are positively correlated with an average a Kendall's tau of 0.76. Given that the best and worst teams are obvious, it is a good sign that Powerwise is also positively correlated with the official rankings with an average tau of 0.63. The gap between tau's is likely attributable to differences in rankings between Powerwise and the RPI for relatively well-matched bubble teams—the object of this analysis.



ii. Discrepancies between historical and Powerwise at-large picks.

Table 1: Powerwise vs. Official At-Large Picks by Season

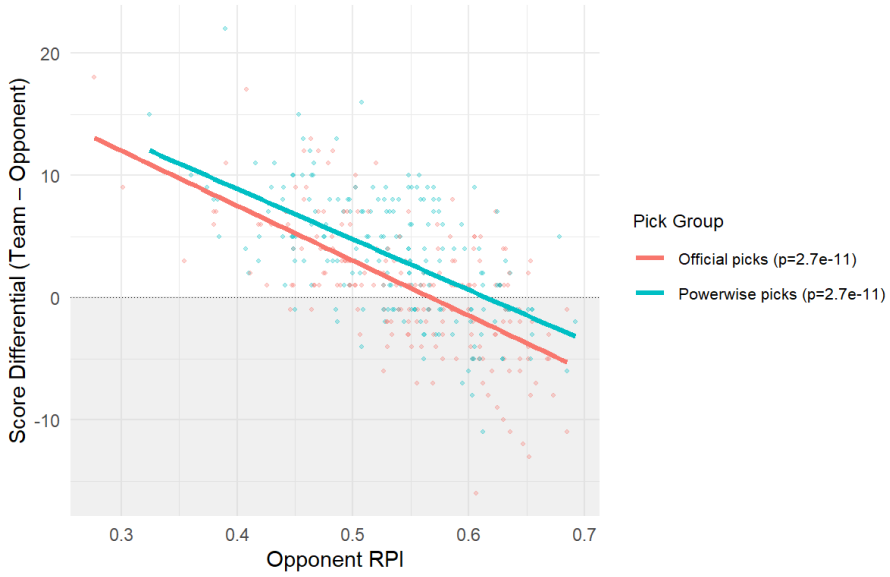
Season	Powerwise picks	Official picks
2022	Notre Dame, Duke	Cornell, Brown
2019	Cornell, Denver	Notre Dame, Johns Hopkins
2018	Rutgers, Penn State, Bucknell	Syracuse, Virginia, Villanova
2017	Duke	North Carolina
2016	Villanova, Stony Brook	Johns Hopkins, Navy
2015	Cornell, Hofstra	Ohio State, Brown
2014	Yale	Harvard

Over the past ten years of data (excluding COVID-19), Powerwise agrees with the official methods' at-large selections in 2013, 2023, and 2024, with only a single discrepant pick in 2014 and 2017. The other years feature additional discrepancies. Figure 1 shows that, as opponent's RPI increases, Powerwise's picks tend to outperform the RPI's picks in terms of per-game score differential during

the regular season. Figure 2 in Appendix C highlights that the opponent strength metric of choice doesn't affect the results. The 2014 figure also contained in Appendix C highlights a case-study in which Yale was far more deserving of an at-large selection than Harvard.

Figure I: Bucketed Score Differential vs Opponent RPI

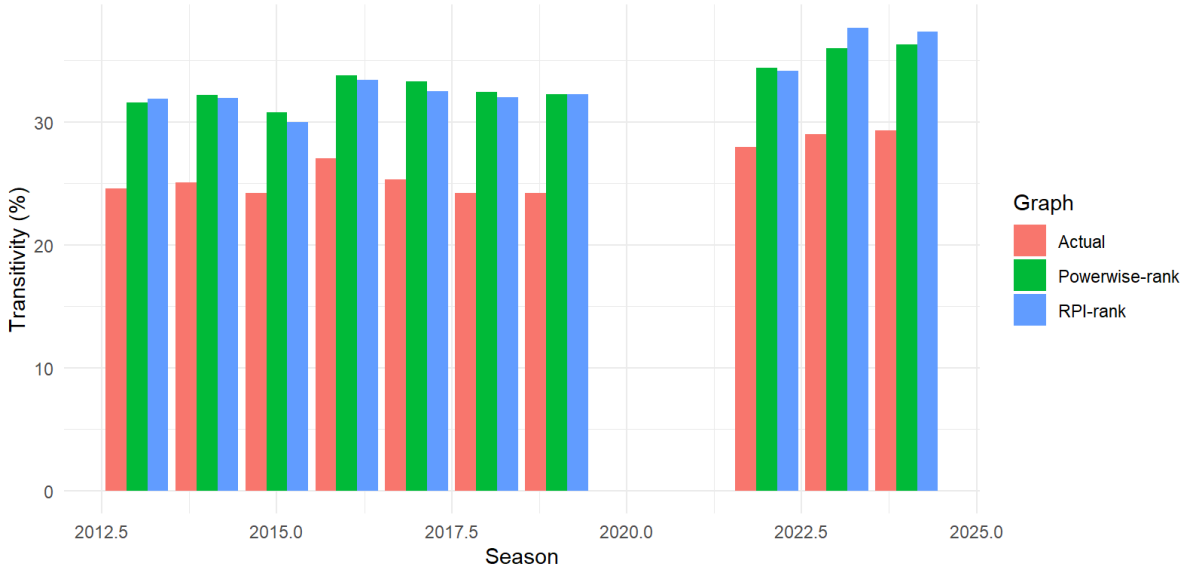
Seasons include: 2022, 2019, 2018, 2017, 2016, 2015, 2014



iii. Transitivity

Transitivity of Win Paths: Actual vs. RPI-rank vs. Powerwise-rank

Pct of A→B→C chains closed by A→C



The above chart shows that both RPI and Powerwise ranking lists better predict game transitivity than the real world. The RPI is ever-so-slightly stricter in recent years, but that's unlikely indistinguishable from randomness. Neither fully captures the "messiness" of real upsets. Whereas

high transitivity shows that picks enforce a strict hierarchy of teams, it doesn't necessarily guarantee accurate rankings—sports involve upsets.

iv. Conference bias

Table 2: Conference Class Effects on At-Large Selection, controlling for Rank

	Dependent variable: 1 = at-large bid, 0 = none	
	(1) Powerwise rank	(2) RPI rank
Intercept	0.367^{***}	0.364^{***}
	(0.068)	(0.067)
Major	$0.112^{\cdot}$	$0.120^{\cdot}$
	(0.067)	(0.066)
Mid-Major	-0.139^{*}	-0.128^{*}
	(0.063)	(0.062)
Small	−0.083	−0.086
	(0.061)	(0.061)
rank.POWERWISE	-0.00568^{***}	
	(0.00062)	
rank.RPI		-0.00572^{***}
		(0.00059)
R^2	0.315	0.323

Standard errors in parentheses. $^{\cdot}p < 0.10$, $^{*}p < 0.05$, $^{***}p < 0.001$.

Table 2 shows that Powerwise does not have a substantially different bias profile to the RPI. Mid-major conferences are most harshly penalized under both methods and, in each case, a one-point worsening in team rank cuts predicted at-large probability by around 0.6%.

Conclusions

Powerwise produces common-sense results with a simple, transparent methodology that emphasizes on-field performance. While Powerwise does not have a substantially different conference-bias profile (over the years selected) to the RPI[2], it does appear to consistently select *more deserving* at-large teams than the current committee-led, RPI-informed approach. Further analysis of ranking stability, time-to-convergence, and predictive power are left to later study.

[2] Additional problems with the RPI are highlighted in Appendix D.

Bibliography

Annis DH and Craig BA (2005) Hybrid Paired Comparison Analysis, with Applications to the Ranking of College Football Teams. *Journal of Quantitative Analysis in Sports* 1(1).

Barrow D, Drayer I, Elliott P, et al. (2013) Ranking rankings: an empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports* 9(2). De Gruyter: 187–202.

Coleman BJ, DuMond JM and Lynch AK (2010) Evidence of Bias in NCAA Tournament Selection and Seeding. *Managerial and Decision Economics* 31(7). Wiley: 431–452.

Colley WN (2002) *Colley's Bias Free College Football Ranking Method:* Princeton University.

Massey K (1997) *Statistical Models Applied to the Rating of Sports Teams*. Bluefield College.

Ochieng PJ, London A and Krész M (2022) A Forward-Looking Approach to Compare Ranking Methods for Sports. *Information* 13(5). 5. Multidisciplinary Digital Publishing Institute: 232.

Paul RJ and Wilson M (2015) Political Correctness, Selection Bias, and the NCAA Basketball Tournament. *Journal of Sports Economics* 16(2): 201–213.

S. I. Gass (1998) Tournaments, Transitivity and Pairwise Comparison Matrices. *The Journal of the Operational Research Society* 49(6): 616–624.

San Diego Union-Tribune (2018) Mountain West announces new host procedure for conference football championship game. Available at: https://www.sandiegouniontribune.com/2018/07/23/mountain-west-announces-new-host-procedure-for-conference-football-championship-game/ (accessed 11 April 2025).

Stefani R (2011) The Methodology of Officially Recognized International Sports Rating Systems. *Journal of Quantitative Analysis in Sports* 7(4). De Gruyter.

Stocks-Smith J (2021) College Basketball Rating (CBR): A new body-of-work metric for NCAA tournament selection. *Journal of Sports Analytics* 7(1). SAGE Publications: 47–55.

USCHO.com | FAQ | Frequently Asked Questions (n.d.) PairWise Rankings explanation. Available at: https://www.uscho.com/faq/pairwise-rankings-explanation/ (accessed 28 May 2024).

Appendix A: The Powerwise Method

A. Simulating matchups for all teams

Powerwise works by simulating a scenario in which all teams in a division "play" against each other in hypothetical pairwise matchups, competing for Powerwise points. For each matchup, one team "wins" and the other "loses" by virtue of a three-step hierarchical procedure that assesses, in order of importance, head-to-head results, each team's record against common opponents, and Power Ratings (to be discussed in the next section). Given a comparison between two hypothetical teams, the Powerwise point is awarded to:

I. The team with the superior **head-to-head record**.
(Continue to step II only in case of a nonexistent or tied head-to-head record.)

II. The team with the superior record against **common opponents**.[3]
(Continue to step III only in case of a nonexistent or tied record against common opponents.)[4]

III. The team with the higher **Power Rating**.

After all matchups are examined, the teams are sorted from most-to-least deserving on the basis of Powerwise points. For MD1 Lacrosse, 76 teams "play" (are compared in) 75 match-ups, so the best possible Powerwise record for a given team is 75-0, and the worst is 0-75.[5]

A notable strength of the Powerwise method is that the majority of its pairwise comparisons are determined by on-field results alone. Analysis of historical lacrosse data (2013-2024) finds that 60–70% of Powerwise points would have been awarded based on a superior head-to-head and common opponents' record.

Table I: Percentage of Powerwise comparisons determined by H2H and CO results

Season	2013	2014	2015	2016	2017	2018	2019	2022	2023	2024	Avg.
% H2H decisive	21.5	19.6	22.2	19.5	18.7	18.9	18.3	18.0	17.6	17.8	∼19%
% CO decisive	72.5	69.4	72.6	67.1	64.7	67.1	64.0	61.8	61.7	60.4	∼66%

This enviable percentage falls as the number of teams in a division rises (assuming a constant number of games per season). As mentioned, when on-field results are indecisive, the Powerwise matchup is determined by an accurate, simple analytic based on goal differentials and featuring an implicit adjustment for strength of schedule described in the next section.

B. Power Ratings

---

[3] Powerwise employs percentage win-loss when calculating common opponent records. One downside of this approach is that it risks downplaying repeated losses. This limitation is discussed in greater detail later.

[4] The suggested Powerwise method bypasses step II in case of a single common opponent game, on the basis that one game is tantamount to randomness. This "ad hoc" adjustment makes the method more complicated to achieve a little more resistance to chance and volatility.

[5] Appendix C, Table 1 presents the 2024 Powerwise ranking list for men's lacrosse.

Powerwise's Power Rating (PR) system is based on goal differentials and designed to estimate the average margin of victory between two teams on a neutral field. The formula a bit "mathematical" for a general audience, but it's simpler than it looks.

Formula 1: Power Ratings

$$\sum_{i=0}^{n}\sum_{j=0}^{m}(PR_i - PR_j) = \sum_{i=0}^{n}\sum_{j=0}^{m}(score_i - score_j) \pm hfa$$

Essentially, calculating PRs involves solving a large system of nonlinear algebraic equations based strictly on game scores and an adjustment for the home-field advantage, iterating until the difference between two teams' Power Ratings is equal to the expected (or real) difference in scores were those two teams to play, adjusting for the home-field advantage.[6]

Power Ratings implicitly account for strength of schedule thanks to the inclusion of goal differentials. To enable the use of goal differentials, PRs employ an adjustment to the score data intended to disincentivize "running up the score." This merits further discussion.

[6] Two options to account for the home-field advantage include a constant adjustment for all teams in a division (determined by historical data or by some other means) or a team-by-team measure that accounts for the fact that different fields (say, Denver's mile-high stadium) perform differently. We opt for the simpler division-by-division average to simplify this analysis, but the preferred method is up for debate.

Appendix B: The Ratings Percentage Index (RPI)

The RPI is a widely-used metric that aims to evaluate a team's strength relative to the strength of its regular season opponents. A formula for the calculation of the RPI is often written as follows, where win-loss records (WL) are equal to wins divided by wins plus losses:

$$RPI = (1/4 * team.WL) + (1/2 * opponents.avg.WL) + (1/4 * opponents.opponents.avg.WL)$$

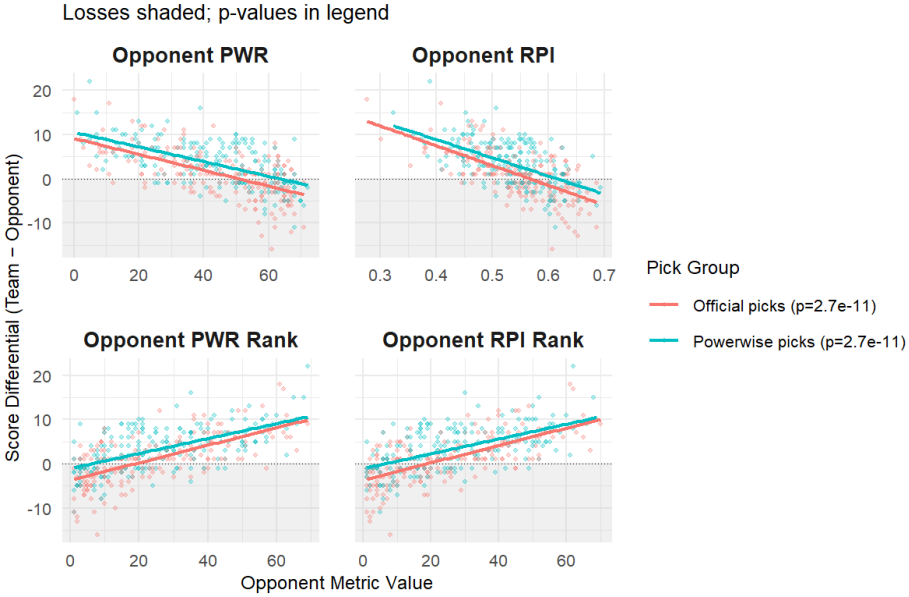
In other words, Team X's RPI is calculated by weighing three components:

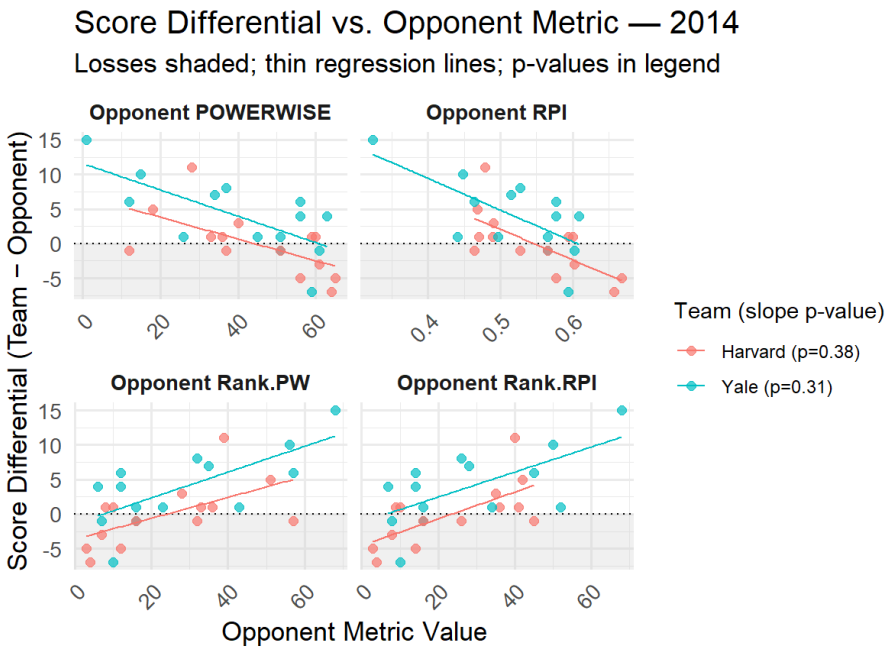
- Team X's win/loss record = 25% of the RPI score.
- Team X's opponents' win/loss record = 50% of the RPI score.
- Team X's opponents' opponents' win/loss record = 25% of the RPI score.

Note that ties are worth half of a win and half of a loss. Wins and losses are often pre-adjusted to account for the home-field advantage, but they're not, currently, in the implementation used for lacrosse.

Appendix C: Additional tables and figures

Figure II: Bucketed Score Differential vs Opponent Metric
Seasons include: 2022, 2019, 2018, 2017, 2016, 2015, 2014





Judging by the 2014 panel, Yale clearly delivered the stronger on-field résumé once you control for opponent strength. Look at the top-left “Opponent POWERWISE” panel: Harvard’s worst defeats are deeper below zero than Yale’s, suggesting Yale was actually more consistent.

In short, Yale outperformed Harvard against the same schedules, with more cushion versus weak teams and fewer drubbings by the strong ones. Powerwise's selection of Yale over Harvard is thus supported by the game-by-game data—even if the committee ultimately went the other way, Yale had the stronger season-long performance.

Appendix D: Three Problems with the Ratings Percentage Index

The Rating Percentage Index (RPI) system for ranking teams in Division I sports, particularly lacrosse, has several flaws that affect its accuracy and reliability. Here are three major deficiencies illustrated by example cases:

Issue 1: Strength of Schedule (SOS) is Based on Opponent Performance, Not Team Performance

The RPI's SOS calculation reflects the win percentage of a team's opponents rather than how a team performed against those opponents. Here's a hypothetical scenario: Should Hampton (Men's D1 lacrosse), ranked 76th in 2024, be invited to the ACC, scheduled tougher competition, and losing all games by a wide margin, despite its losses, Hampton's RPI would jump from 0.3179 to 0.5053, improving by 38 places due to the stronger schedule alone.

This suggests the RPI is more sensitive to a team's opponents' performance rather than its own, distorting rankings. The formula's reliance on opponent win percentages (OWP) and opponents' opponents win percentages (OOWP) creates a situation wherein a team can artificially rise in the rankings by merely playing tougher competition, regardless of actual performance. Table B1 presents the actual RPI results from the 2024 season. If instead, Hampton was to have played in the ACC with a different schedule, table B2 shows Hampton's RPI improvement in rating and rank, from 0.3179 to 0.5053 or 76th to 38th, despite remaining the same team.

RPI Rank	Team Name	RPI	Wins	Losses
70	Mt. St. Mary's	0.3663	1	14
71	St. Bonaventure	0.3538	1	11
72	Wagner	0.3468	1	12
73	Queens	0.3353	2	11
74	Mass-Lowell	0.3350	0	12
75	Lindenwood	0.3235	0	12
76	**Hampton**	**0.3179**	**0**	**13**

RPI Rank	Team Name	RPI	Wins	Losses
38	**Hampton**	**0.5053**	**0**	**13**
39	Vermont	0.4907	8	8
40	Drexel	0.4883	5	9
41	Air Force	0.4876	9	6
42	Quinnipiac	0.4874	9	5
43	Brown	0.4790	3	11
44	LIU	0.4730	10	4

Tables D1 and D2: Hampton Dilemma. Hampton ranked 76th during the 2024 season. When the schedule was adjusted so that Hampton played in the ACC, the same team with a more difficult schedule had an RPI ranking of 38.

Issue 2: Ignoring Goal Differentials Invites Inaccuracy

The second major issue is that RPI only accounts for wins and losses, disregarding the margin of victory or defeat. Consider Team A (which defeats teams by large margins) versus Team B (which wins against the same teams by narrow margins). The RPI system would rate both teams similarly, despite Team A's superior performance. This overlooks the fact that a decisive win (e.g., 15-0) indicates a stronger team than a close win (e.g., 13-12). The absence of goal differential data reduces the granularity of the analysis and leads to less accurate rankings. This approach is mathematically questionable, as it ignores key performance metrics that better reflect a team's strength.

Issue 3: Hypersensitivity to Irrelevant Games

The RPI's sensitivity to small changes in irrelevant or inconsequential games further compromises its accuracy. The Lafayette vs. Delaware example illustrated in Table D3 shows how a remote upset game, which has little bearing on tournament qualification, can significantly impact the rankings.

The right panel of D3 shows a shift in the rankings of multiple top-20 teams, despite Lafayette and Delaware having little direct interaction with the tournament contenders. Such hypersensitivity undermines the sense that teams control their own destiny.

Table D4 repeats this experiment with Power Ratings to show that they're less sensitive to similar perturbations in the data. Only one, explainable, ranking change occurs in the left panel of D4.

Rank	Team	RPI		Rank	Team	RPI
1	Notre Dame	0.7100		1	Notre Dame	0.7097
2	Duke	0.6632		2	Duke	0.6631
3	Johns Hopkins	0.6520		3	Johns Hopkins	0.6517
4	Syracuse	0.6404		4	Syracuse	0.6381
5	Virginia	0.6371		5	Virginia	0.6368
6	Denver	0.6246		6	Denver	0.6246
7	Maryland	0.6223		7	Maryland	0.6223
8	Princeton	0.6158		8	Penn State	0.6157
9	Penn State	0.6158		9	Princeton	0.6157
10	Georgetown	0.6147		10	Georgetown	0.6146
11	Penn	0.6044		11	Cornell	0.6038
12	Cornell	0.6041		12	Yale	0.6018
13	Michigan	0.6019		13	Penn	0.6018
14	Yale	0.6017		14	Michigan	0.5994
15	St Joesphs	0.5970		15	St Josephs	0.5967

Table D3: RPI ratings are hypersensitive. Left presents the same data as right, except for switching the result of a close, remote game played early in the season between relatively irrelevant teams, Delaware and Lafayette. Rankings for teams 8, 9, 11, 12, 13, and 14 were displaced. RPI ratings are so hypersensitive that the difference between Princeton and Penn State occurs after the ten thousandth decimal place.

Rank	Team	PR		Rank	Team	PR
1	Notre Dame	99.90		1	Notre Dame	99.90
2	Duke	97.51		2	Duke	97.52
3	Virginia	97.38		3	Virginia	97.37
4	Syracuse	97.02		4	Syracuse	97.00
5	Penn State	96.90		5	Penn State	96.91
6	Johns Hopkins	96.77		6	Johns Hopkins	96.77
7	Princeton	96.46		7	Princeton	96.47
8	Georgetown	95.78		8	Georgetown	95.78
9	Maryland	95.67		9	Maryland	95.68
10	Cornell	95.67		10	Cornell	95.68
11	Yale	95.47		11	Yale	95.49
12	Denver	95.47		12	Denver	95.47
13	Michigan	95.25		13	Michigan	95.21
14	Towson	94.80		14	Army	94.82
15	Army	94.68		15	Towson	94.69

Table D4: The same Delaware vs. Lafayette game had very little impact on the teams' power ratings. Only Army and Towson switch places (because Lafayette's improvement in the power ratings hurts Army—an opponent of Lafayette).

In conclusion: These issues with RPI—its flawed SOS calculation, its disregard for goal differentials, and its hypersensitivity to irrelevant games—compromise its effectiveness as a ranking system. As a result, the RPI fails to provide a true reflection of a team's ability, performance, and undermines the sense that teams control their own destiny. This evidence highlights the need for accurate and sophisticated systems that can incorporate additional data, such as goal differentials, and reduce the impact of arbitrary results. Power ratings better represent a team's competitive standing.