Equitable Divisional Alignment for High School Cross Country Competitions

Matt Kretchmar
Department of Mathematics and Computer Science
Denison University
Granville, OH 43023
Email: kretchmar@denison.edu

***Abstract*—Ohio's process of sorting school districts into different athletic divisions based on school enrollment is shown to create inequities in the sport of cross country running. The nature of order statistics greatly amplifies how differences in school enrollment induce differences in competitive results. In this work, we investigate properties of the division process which have an effect on competitive equity. We then explore alternative solutions to the division creation process and study their potential for promoting fair competition.**

I. Introduction

Our goal is to study how to best partition Ohio's high schools into different athletic divisions to promote fair competition in the sport of cross country racing (XC). The wide enrollment variation among Ohio's high schools creates divisions where very large schools compete against much smaller schools. In our prior work, we revealed how this size discrepancy translates into inequity in competition for winning a state championship title[1]. Our goal here is to build on that prior work by studying a number of alternative processes for creating athletic divisions and comparing their competitive equity to the status quo.

We use statistical modeling and Monte Carlo simulation as our primary tools of investigation. A statistical model of runners' performance is used to create virtual runners and teams for Ohio's high schools. We simulate these teams competing against each other using Monte Carlo techniques. The simulated XC seasons produce statistical records of wins and losses for each team. We use a metric called the normalized win rate to examine how each team fares in these simulated competitions.

A. Map of the Paper

We continue this section with an overview of the division-creation problem in the state of Ohio, and we discuss how similar modeling techniques have been used previously in sports analytics. Section II gives the details of our statistical model and how those are employed in Monte Carlo simulations. We also define the normalized win rate as a key metric for equity. Section III presents several alternative methods of creating athletic divisions. In Section IV we look at different properties of these alternative methods to see how they impact the normalized win rate metric. In Section V we analyze the results of different methods of divisional alignment to determine which methods promote greater equity. Section VI summarizes our main findings and points to possible future work.

B. Background

The Ohio High School Athletic Association (OHSAA) governs and oversees Ohio's high school interscholastic athletic competitions[6]. Part of their responsibility is to sort Ohio's 735 recognized high school programs into different divisions in order to create fair competition for the teams involved. In the sport of cross country (XC), OHSAA currently divides the schools into three different, equally-sized athletic divisions[7]. The intention of the division process is to create an equal playing field of three divisions with the same number of teams, thus ensuring no team is advantaged or disadvantaged by competing against a greater or lesser number of teams. This division process is done independently for boys and girls cross country programs; in this paper we concentrate only on girls XC competition assuming that the analysis of boys teams is similar. We define a *division process* as a method of partitioning XC teams into different athletic divisions; our goal in this paper is to examine multiple alternative division processes and to understand their properties.

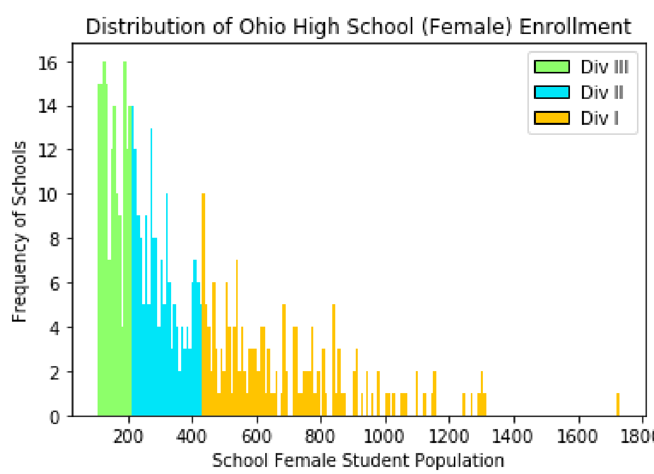

Fig. 1. School Enrollment Distribution

Figure 1 shows the distribution of school population (female student enrollment numbers) for the 501 high schools in the state of Ohio that fielded a cross country team in 2018[8]. Also shown are the three XC divisions (Div I in orange, Div II in blue, Div III in green). The difficulty identified in our earlier work is that there is still a large variation in school population within each division, especially in Division I which contains a small handful of very large schools. This size discrepancy creates significant inequity in competition.

C. Prior Work

Both statistical modeling and Monte Carlo simulation have an extensive and long history of application to sports analytics. Even as early as 1974, researchers were using Monte Carlo simulation to study the effects of batting order choice in baseball [11]. Other researchers have applied similar simulations to measure outcomes [9] and duration [10] in tennis matches. Certainly the explosion of Sabermetrics following the popularity of *Money Ball* shows just how wide sports analytics has spread and also entered the lexicon of popular culture [12]. We follow in this same tradition to apply similar techniques to distance running performance.

II. Model and Equity Metrics

Here we briefly review the model and equity metrics we use to simulate competition and measure the fairness of various division creation processes. A more detailed presentation of these same concepts is available in our prior work[1].

A. Model

We use data from HeathLine[4] and Strava[14] to model the 5k per-mile pace of 14-18 year old females; the distribution of which is shown in Figure 2. We fit a beta model to this distribution as shown by the orange line. For a school with N female students, we draw N samples from this distribution to form the school's entire female student population. We then select the top seven samples to form that school's girls XC team.

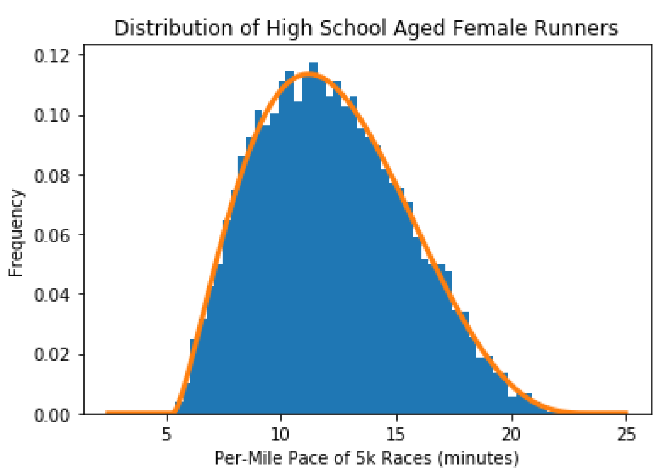

Fig. 2. Distribution of Per-Mile Pace of Female High School Students

We use a different beta distribution to model the *variation* in running pace for each runner at each school. By sampling from these distributions (one model per runner), we obtain the running pace and race finish time for each runner on a given day of competition. Once we have each individual runner's 5k time on a given day, we rank the runners by finish time, compute each team's score, and determine the winner of each meet[3]. In some experiments we simulate a meet with a small number of teams to study the effect of one parameter in the model. In other cases we simulate Ohio's end-of-season process leading to a state championship:

$$\text{District Race} \rightarrow \text{Regional Race} \rightarrow \text{State Meet}$$

B. Metric

Our goal is to study how competitive equity for a state title is impacted by the process of creating athletic divisions. The standard we established in our earlier work is that of proportionality[2]: a school's winning rate (of the state title) should be in proportion to its population (female enrollment) within its division. To evaluate winning percentage on the basis of proportionality we create a normalized win rate. A team's actual winning rate is computed as the number of simulated championships divided by the total number of simulated seasons. This win rate should be in proportion to the school's relative size. We compute the expected win rate by measuring the school's size as a fraction of all Ohio students (sum of all school sizes in that division).

$$\begin{aligned} \text{normalized win rate} \qquad &= \frac{\text{actual win rate}}{\text{expected win rate}} \\ &= \frac{\frac{\text{championships}}{\text{seasons}}}{\frac{\text{school population}}{\text{total Ohio student population}}} \end{aligned}$$

In a fair system, each school should have a normalized win rate of approximately 1.0; the actual win rate from the simulation should be equal to the expected win rate based on enrollment. Figure 3 shows the normalized win rate for all three Ohio divisions using the current division process. The y-axis is in logarithmic scale to more accurately reflect the low winning rate for smaller teams. Schools are plotted with a green dot if their normalized win rate is above 1.0, and a red dot if the normalized win rate falls below 1.0. In Division I, the smallest school has a normalized win rate of about 0.054 (winning about 18.5 times less often than they should, while the largest school has a normalized win rate of 5.75 (winning nearly 6 times more often than they should).

While we can use normalized win rate as the metric to compare the current division process with proposed alternatives, we must set some guidelines for interpreting normalized win rates in terms of equity. The ideally division process would produce a normalized win rate of 1.0 for each team, meaning that a team wins the state title at a rate proportional to their school population. This will not be achievable in practice, but it can serve as an inspirational goal. We settle for applying a scale to interpret the acceptability of normalized win rates.

We say a team is being treated *acceptably* if its normalized win rate falls between 0.67 and 1.5. This implies a team is not winning 50% more or less often than it should on the basis

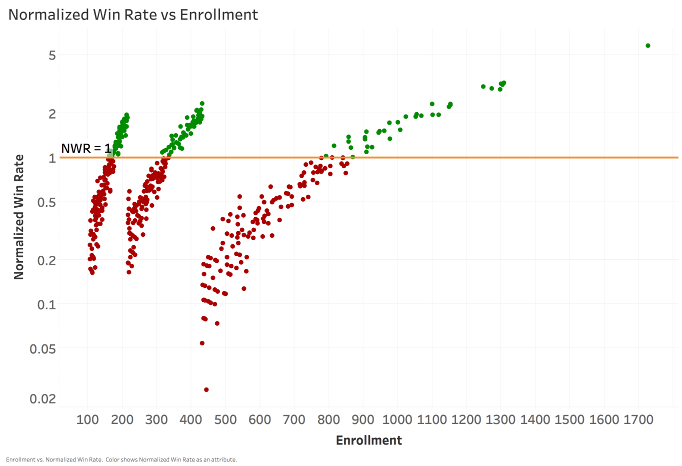

Fig. 3. Normalized Wins for All Teams

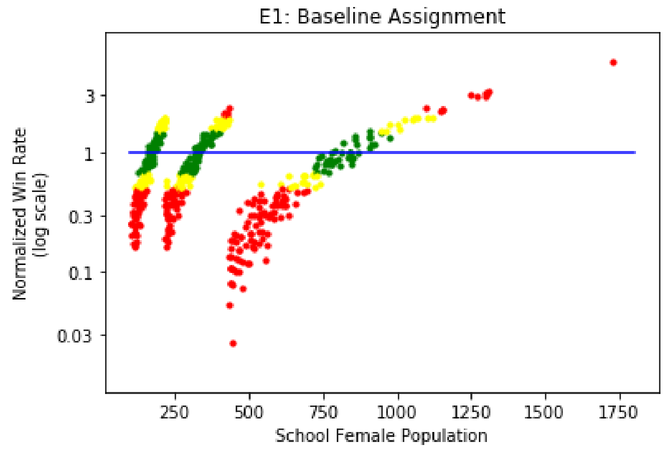

Fig. 5. E1: Normalized Win Rate

of the school's population. This is a fairly "generous" range for acceptability. A team is treated *marginally* if its win rate is between 1.5 and 2.0 (winning up to twice as often) or if the win rate is between 0.67 and 0.5 (winning up to half as often). A team is treated *unacceptably* if its normalized win rate is above 2.0 or below 0.5 – winning or losing at more than twice the rate as it should.

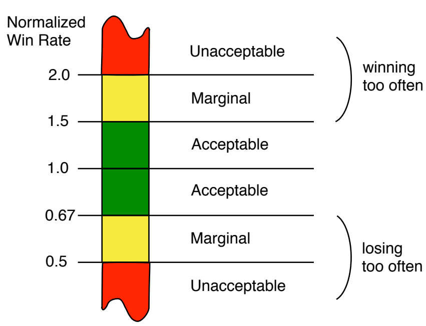

Fig. 4. Scale of Fair Normalized Win Rate

This scale is now applied using the current division process to re-plot the XC teams in Figure 5. The teams now plotted in red have unacceptable normalized win rates (win way too often or way too infrequently), teams in yellow are marginal, and teams in green have acceptable normalized win rates.

As we consider alternative division processes, we use the normalized win rate to compute the following measurements. For each alternative division process, we conduct a full simulation (typically 10,000 seasons) to compute the normalized win rate for each school using each division process. Applying the scale we count

- **Number Acceptable**: The number of XC teams with an acceptable normalized win rate.
- **Number Marginal**: The number of XC teams with a marginal normalized win rate.
- **Number Unacceptable**: The number of XC teams with an unacceptable normalized win rate.

Obviously we seek a division method which maximizes the number of acceptable teams and minimizes the number of marginal and especially unacceptable teams. We will also measure

- **Max NWR**: The maximum normalized win rate for any XC team.
- **Min NWR**: The minimum normalized win rate for any XC team.
- **Average NWR**: The average normalized win rate of XC teams[1].

$$e^{\left(\frac{1}{n}\sum_{i=1}^{n}|\log(\mathrm{nwr}_i)|\right)}$$

The maximum NWR tells us how much the best team is winning, while the minimum shows how infrequently the worst team is winning. The average NWR metric is especially helpful in that it gives us a feel for deviation of the average from the ideal 1.0 normalized win rate for the whole collection of teams.

III. Alternative Division Processes

The current division process creates three divisions with the same number of schools in each division. The largest third of Ohio schools are in Division I, the middle third in Division II and the smallest third in Division III. Presumably, the goal is to create equal sized divisions so that no school has to compete against more or fewer teams. In this section we introduce other possible division processes, many of which we will subsequently analyze using the normalized win rate metrics.

While there are a multitude of potential division processes, we will only consider those that are based on a break point type of method. For these break point methods, we align all XC teams in a row, ordered from the smallest school to the largest school and set *break points* that mark the boundary between divisions. Figure 6 illustrates the break points for the

[1]This is computed as an average of the absolute value of the $\log(nwr)$ so that both low and high NWRs can be averaged together appropriately.

current division process where the different divisions are color coded. Other division processes differ by where and how they establish break points.

Note that this class of methods (the break point methods) create divisions where there is no overlap in size/enrollment; all the schools in the big division are larger than all the schools in the next division, and so on. In keeping with Ohio traditional nomenclature, we call "Division I" the division which contains the largest schools (right-most in Figure 6). Division II will come next and so on, down to the last division comprised of the smallest schools (Division III or Division IV in our examples).

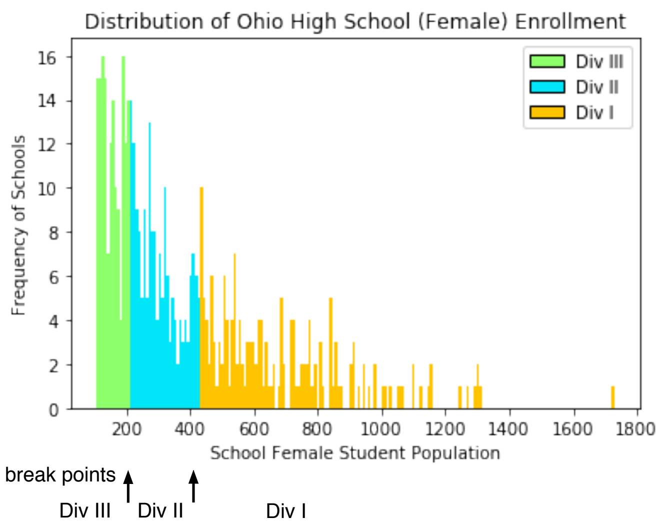

Fig. 6. Illustration of Breakpoints

A. Number of Divisions

An obvious way to modify the division process is to change the number of divisions. Not all Ohio high school sports have the same number of divisions. Football has seven divisions. Volleyball, basketball and baseball each have four, while other sports have only one or two divisions. Football would have four divisions if it were to maintain something close to the same number of teams in each division as does cross country. Though OSHAA does not publish its deliberations, we hypothesize that the number of divisions chosen for each sport depends on the number of schools competing in that sport, the logistics of competition, and the cost of managing those divisions. Our prior analysis shows that the inequity in the sport of XC arises from a large difference between the smallest and largest school within each division, especially in Division I. Increasing the number of divisions would reduce the enrollment disparity within each division. We explore division processes with three and four divisions.

B. Team Caps

Because the distribution of school populations is exponential, much of the inequity arises from a small handful of very large schools. Many of Ohio's largest school districts have chosen to divide themselves into two or more separate high schools as they grow. For example, Hilliard has divided into three separate high schools (each of which still ranks among the largest schools in Division I), and Olentangy has divided into four high schools (again, each of which is significantly large compared to Div I averages). Conversely, Mason has remained as a single high school with an enrollment of 3500 students, eclipsing all other programs in Division I by a significant margin. One solution is to introduce a cap on the maximum school size. For example, to implement a cap of 1000 students, schools that have more than 1000 students will create two separate XC teams. These teams can share coaches, workouts, facilities, etc. However, for the purposes of competition, they divide and compete as separate teams. We can imagine that newly entering runners are randomly placed into one of the two sub-teams; this ensures that teams retain their basic rosters from season to season. We study models with and without a cap of 1000 students.

C. Different Divisional Methods

Another tactic is to place schools in divisions differently, to set the break points between divisions differently. The current strategy aims for equity by keeping each division the same size (same number of teams). The break points between divisions are based on dividing the total count of schools into thirds. But there are other ways to choose break points. One option is to choose break points that retain the same absolute size difference in each division. Another is to choose break points for the same relative size difference in each division.

For descriptive purposes, let's assume that N_1 is the population of the smallest school in a division and N_2 is the population of the largest in that division.

- **Absolute Size Difference** (absolute difference) is measured as the difference between the largest and smallest enrollments. $N_2 - N_1$.
- **Relative Size Difference** (relative difference) is measured as the relative difference (ratio) between the largest and smallest enrollments. $\frac{N_2}{N_1}$.

The table below shows the absolute and relative differences for the current divisional system[2]. Note that by keeping the same number of teams in each division, the absolute and relative differences for each division vary widely.

Division	Largest N_2	Smallest N_1	Absolute Difference	Relative Difference
I	1728	433	1295	3.99
II	432	217	215	1.99
III	216	106	110	2.03

Fig. 7. Absolute and Relative Difference in Current System

To create divisions with the **same absolute difference**, we divide the overall range by 3 to find the interval or absolute difference for each division.

[2]The exact largest and smallest are not published for Division II and Division III, so the authors had to estimate them from public data.

$$
\begin{aligned}
\text{range} &= \text{largest} - \text{smallest} = 1728 - 106 = 1622 \\
\text{interval} &= \frac{\text{range}}{3} = \frac{1622}{3} = 540.67 \\
\text{break point 1} &= \text{smallest} + \text{interval} \\
&= 106 + 540.67 = 646.67 \\
\text{break point 2} &= \text{smallest} + 2 \cdot \text{interval} \\
&= 106 + 2 \cdot 540.67 = 1187.33
\end{aligned}
$$

Division	Largest	Smallest	Count
I	1728	1249	7
II	1154	648	73
III	640	106	424

Fig. 8. Divisions with Equal Absolute Difference

The resulting partition for the Equal Absolute Difference process is shown in Table 8. The obvious problem is that Div I now has 7 schools while Div III has 424. The highly skewed distribution of school enrollments makes this potential division process highly problematic.

To create divisions with the **same relative difference**, we compute the cube root of the ratio between the overall largest and smallest schools.

$$
\begin{aligned}
\text{factor} &= \left(\frac{\text{largest}}{\text{smallest}}\right)^{\frac{1}{3}} = \left(\frac{1728}{106}\right)^{\frac{1}{3}} = 2.536 \\
\text{break point 1} &= \text{smallest} \cdot \text{factor} = 268.8 \\
\text{break point 2} &= \text{smallest} \cdot \text{factor}^2 = 681.7
\end{aligned}
$$

Division	Largest	Smallest	Count
I	1728	682	75
II	681	269	206
III	268	106	221

Fig. 9. Divisions with Equal Relative Difference

This process creates a Div II and Div III with approximately equal size, but with a smaller Div I. However, it is not as imbalanced as the same-absolute-difference process.

In Figure 10 we see a graphic of all schools competing in XC and how the different division processes would allocate teams to divisions. The current method, with equal number of teams in each division, shows just how large the spread is in Div I (orange) and how small it is comparatively in Div II and Div III. The method of equal absolute difference captures only the small handful of large teams in Div I while creating large numbers of teams in Div II and especially Div III. The method of equal relative difference seems to find a happy medium between the two extremes.

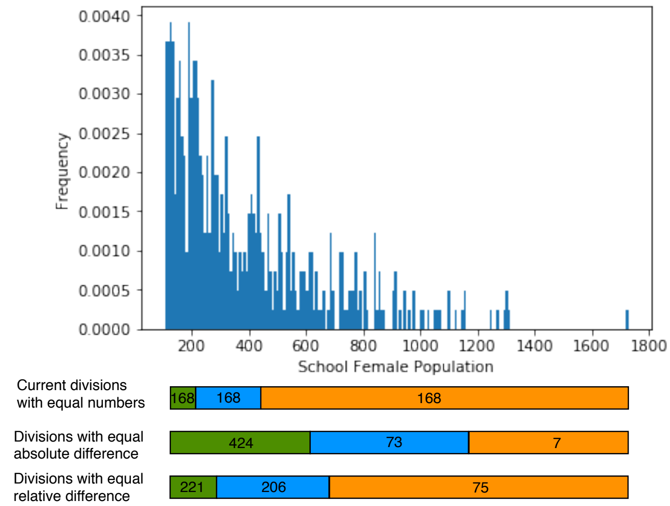

Fig. 10. Alternate Methods of Divisional Alignment

IV. Sensitivity Analysis

Before we embark upon researching the equity of potential division processes, we first examine how various properties of division processes affect normalized win rate. We study the *sensativity* of the normalized win rate to various factors to give us insight as to which potential division processes might be most equitable. We explore the following three characteristics of processes:

- **Absolute Difference**: We study the effect of varying the absolute difference of two school populations on their normalized win rates in simulated meets. We keep the relative difference fixed.
- **Relative Difference**: Similarly we keep the absolute difference of two school populations fixed and vary their relative difference to determine the sensitivity of normalized win rates.
- **Independence of Irrelevant Alternatives**: Lastly we investigate the sensitivity of the normalized win rate to the number of schools competing in a meet.

In the first two sets of experiments, we conduct a series of simulated meets of 10 schools. Their populations are denoted $N_1 < N_2 < N_3 < ... < N_{10}$. In general, we will be most interested in examining the normalized win rate of the smallest and largest schools to see how they are affected by variation in the absolute or relative population differences. We will keep the school populations uniformly distributed between N_1 and N_{10}, that is $N_i = p \cdot (i-1) + N_1$ where p is the difference in population between two consecutive schools.

A. Sensitivity to Absolute Difference

In this first experiment we keep the relative difference in population of the two target schools fixed at a constant: $\frac{N_{10}}{N_1} = 2$. We vary $N_1 \in \{100, 200, 300, 500, 1000, 1500, 2000\}$ and set N_{10} as twice the size of N_1 to keep their relative difference fixed. The other eight schools are distributed between N_1 and

N_{10}. Obviously, the absolute difference between N_1 and N_{10} will change. We simulate 10,000 independent seasons with each set of population values (70,000 simulated seasons in all). For each season, we resample each school's entire female population, select the top seven for the XC team, and then have each team compete in a single 10-team meet. We compute the normalized win rate for each school.

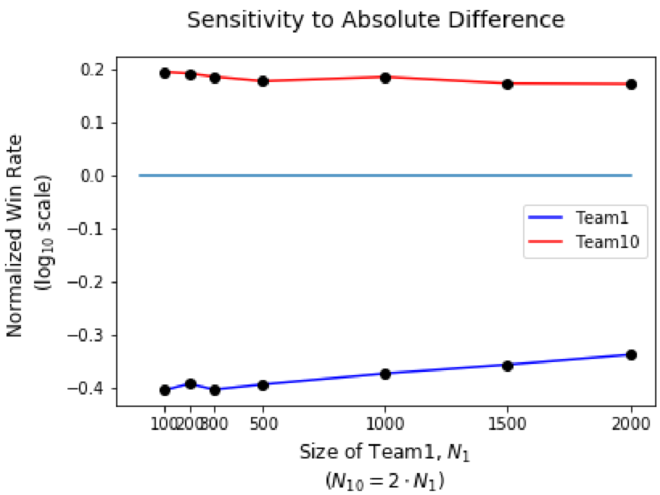

Fig. 11. Sensitivity to Absolute Difference in Population

Figure 11 shows the results of the simulations with these ten teams. At the smallest population sizes (when $N_1 = 100$, and $N_{10} = 200$) we see that Team1 has a normalized win rate of $10^{-0.4} = 0.398$ indicating that the small school is winning this meet about $\frac{1}{0.398} = 2.5$ times less often than they should. Conversely Team10 has a normalized win rate of $10^{0.2} = 1.58$ indicating they are winning almost 60% more often than they should. As we increase the population of both schools (keeping the ratio of $\frac{N_{10}}{N_1} = 2$), we see that the normalized win rates of both schools do not change substantially. Team10's normalized win rate is almost flat, while Team1's normalized win rate rises slightly.

We conclude that normalized win rate is relatively insensitive to the absolute difference in school populations. It does not seem to matter if the absolute different between Team1 and Team10 is 100 or 2000; their normalized win rates do not differ dramatically as long as the relative difference between them is fixed.

B. Sensitivity to Relative Difference

We repeat the same set of experiments except this time we keep the absolute difference of the two schoos fixed: $N_{10} - N_1 = 200$. We again vary the population of the small school over the same set of values ($N_1 \in \{100, 200, 300, 500, 1000, 1500, 2000\}$) and set $N_{10} = N_1 + 200$. The relative difference $\frac{N_{10}}{N_1}$ between the two schools will vary from 3 to 1 as we change the population. Again, the other eight schools are distributed between N_1 and N_{10}. We repeat the simulation of 10,000 independent seasons (10,000 meets) for each set of population values.

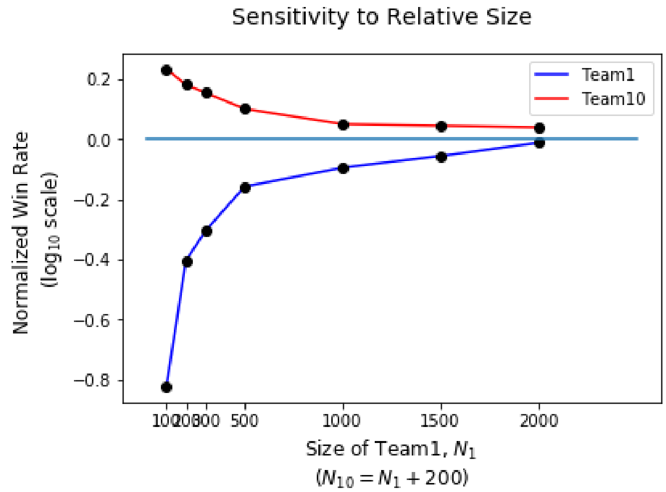

Fig. 12. Sensitivity to Relative Difference in Population

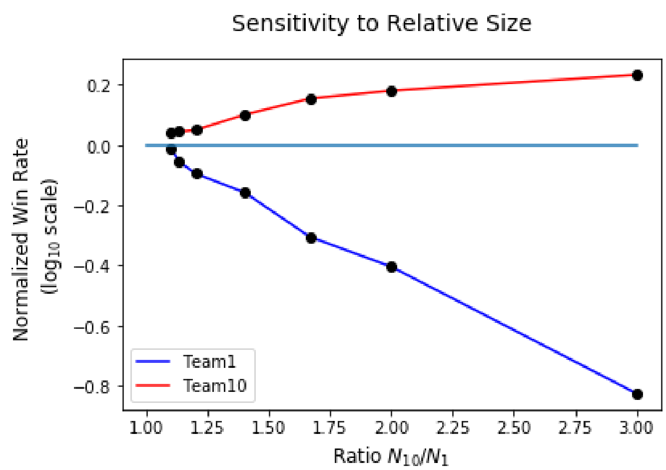

Fig. 13. Sensitivity to Relative Difference in Population

Figure 12 shows how the normalized win rates change as we vary population. Figure 13 shows the same data where the x-axis is now the relative difference between Team1 and Team10, more clearly demonstrating the sensitivity of the normalized win rates to changes in relative population size. As the relative difference approaches 1 (when both schools are almost the same size), the normalized win rates for both programs hover near the desirable mark of 1.0. As the ratio between the schools increases, we see a dramatic separation in normalized win rates. The normalized win rate for Team10 grows steadily, ultimately rising to a place where Team10 wins 1.7 times more often than warranted. The normalized win rate for Team1 falls more dramatically, reaching a point where Team1 is winning 6.3 times less often than they should.

We conclude that normalized win rate is highly sensitive to changes in relative population difference, especially for the smaller schools in the competition.

C. Sensitivity to Independent Alternatives

In this final set of experiments, we change the number of teams in the competition. We keep $N_1 = 500$ and $N_{10} = 2000$ so that the relative and absolute differences in school populations are fixed. We vary how many *other* teams are included in the competition. We change the number of teams from 2 (a dual meet between Team1 and Team10) to 20 teams. Again, all other team populations are evenly distributed between the smallest school (at $N_1 = 500$) and the largest school ($N_{10} = 2000$).

Figure 14 shows the sensitivity of normalized win rates to the number of schools competing in the meet. When there are just two teams (a dual meet), Team10 has a normalized win rate of 1.20 and Team1 has 0.19. This is already not a fair situation, but it gets only worse as more teams are added. The normalized win rate of Team1 decreases dramatically as more teams are added, falling to a value of 0.065 with a 20-team competition; this is winning more than 15 times less often than they should. The interesting thing to note is that Team10's normalized win rate *increases* as more teams are added. While their number of wins decreases, their winning rate increases substantially beyond it's desired 1.0 value based on proportional population. Their normalized win rate starts at 1.20 and rises to 1.85 (winning almost twice as often as they should).

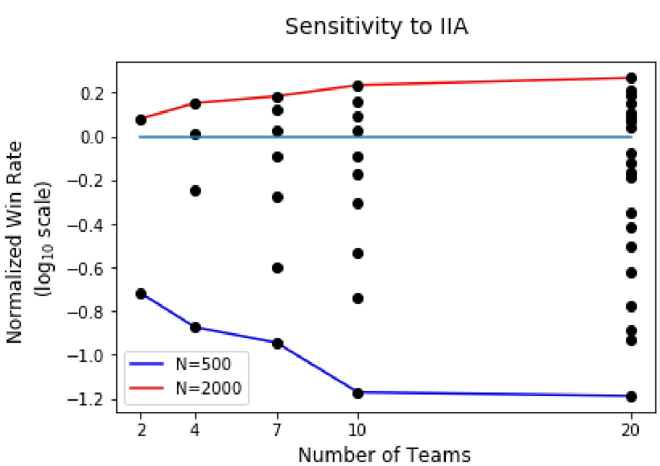

Fig. 14. Sensitivity to Independence of Irrelevant Alternatives

We call this type of sensitivity *independence of irrelevant alternatives* because it carries the same feel from the statistical term. Adding more teams will shrink the win count of both teams, but their ratio of wins should remain static; it clearly does not. We hypothesize the cause of this effect. When there are just two teams, it is unlikely but possible that the smaller Team1 beats the larger Team10 in a competition; it could happen from time to time. As we add more teams, Team1 now needs to independently beat a larger number of bigger teams. They must have multiple unlikely scenarios all arise at the same time. Team10, meanwhile, wins more often possibly because the scoring runners from potential upsetting teams are getting pushed further and further down the finishing rankings as more teams are added to the mix. This makes it less likely that a smaller team upsets a bigger team.

We conclude that normalized win rate is moderately sensitive to the number of other schools in the competition. This is especially important for the state championship scenario as the smaller teams must have upsets at the district meet, the regional meet and then again at the 20-team state meet. This makes an upset bid by small teams statistically improbable.

V. Analysis

In this section, we analyze several potential alternative methods of divisional alignment using the normalized win rate metric. We start by explicitly identifying the solutions/methods we investigate. There are three primary methods of choosing break points for the division process:

- **Method of Equal Numbers** attempts to keep the same number of teams in each division.
- **Method of Equal Absolute Difference** keeps the same absolute different in population ($N_{\text{large}} - N_{\text{small}}$) for each division.
- **Method of Equal Relative Difference** keeps the same relative difference in population $\left(\frac{N_{\text{large}}}{N_{\text{small}}}\right)$ for each division.

For each of these methods of choosing break points, we explore their application with both three divisions (current method) and four divisions (adding one more). That gives us six combinations. We then also implement a cap system whereby we set an upper limit on school size; schools in excess of the cap must form two distinct teams where their runners are randomly assigned to each team. We try a cap of $C = 1000$. This gives us 12 different permutations for 12 different experiments. We label the experiments E1 through E12 as indicated in Figure 15. For each experiment, we run 10,000 simulated seasons and compute the normalized win rate for each school[3]. Experiment E1 is the current division process used by OSHAA; E2 through E12 indicate possible alternatives.

	Cap	3 Divisions	4 Divisions
Method of Equal Numbers	none	E1	E2
	1000	E3	E4
Method of Equal Absolute Differences	none	E5	E6
	1000	E7	E8
Method of Equal Relative Differences	none	E9	E10
	1000	E11	E12

Fig. 15. Table of Experiments

[3]Because each division process produces divisions of different sizes, we modify the number of districts and regions in each scenario. The goal is to keep 12-14 teams in each District, choosing the top 5 to move on Regions. We then alter the number of teams moving on from each Regional championship to keep a state meet of 15 to 20 teams.

As noted in the section below, we were unable to complete four the experiments. E5-E8 all stem from the Method of Equal Absolute Difference. Recall the size of the divisions created using this method varied dramatically. Thus they are not practical for implementation or for this study.

The graphs and tabulated data for the remaining eight experiments are shown in the appendix where they can be displayed on one page to facilitate visual comparison. Figure 25 shows the plots of normalized win rates for each school using each different division process. Figure 24 gives the complete tabular results for the eight experiments. We place them in an appendix so that all eight may fit on one page and be examined together. In the remainder of this section, we investigate different trends observed in the eight experiments.

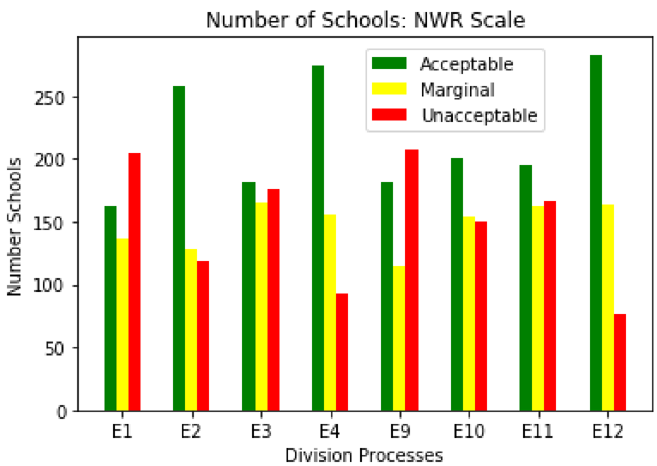

Fig. 16. Experiments: Number of Schools Acceptable, Marginal, Unacceptable

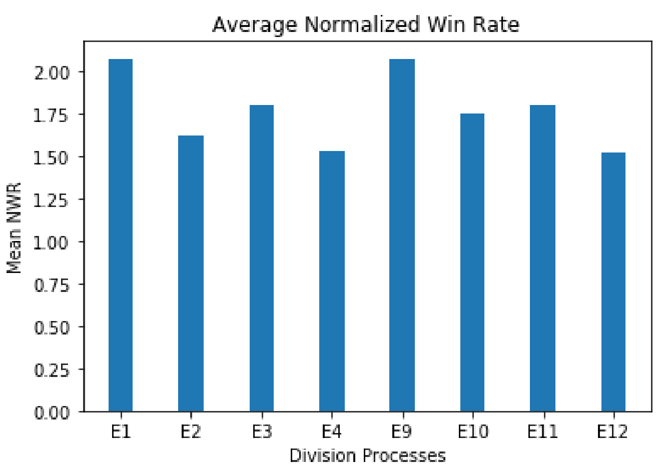

Fig. 17. Experiments: Mean Deviation from Ideal NWR=1

A. E5-E8: Methods of Equal Absolute Difference

We first discuss the four experiments of Equal Absolute Difference. As alluded to when we introduced this break point method, the main difficulty is that it produces divisions with extreme imbalances in team numbers. Division I is comprised of only a small handful of schools, while Division III is comprised of about three-quarters of all schools. We consider these solutions to be impractical and thus do not simulate them for their equity. There are no graphical results or tabulated data for these four methods.

B. Number of Divisions

Among the eight remaining experiments, four feature 3 divisions and four have 4 divisions. In each case, increasing the number of divisions from 3 to 4 had a significant improvement on competitive equity. These results are averaged in the table in Figure 18. We can see that as we increase the number of divisions from 3 to 4, the number of schools with an acceptable normalized win rate increases from 180 to 254, and the number of schools with unacceptable NWRs drop by nearly 80. These counts are shown graphically in Figure 19.

Num Div	Num Accept	Number Marg	Number Unaccept	Normalized Win Rate
3	180.25	144.50	188.75	1.94025
4	254.00	150.25	109.25	1.60950

Fig. 18. Average Results, 3 vs 4 Divisions

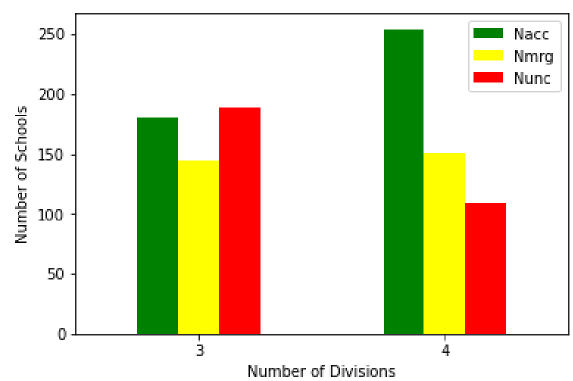

Fig. 19. 3 vs 4 Divisions: Acceptable/Marginal/Unacceptable

The average normalized win rate also improves substantially. The average NWR drops from 1.94 to 1.61 by adding another division. While both marks are ”marginal”, the addition of a fourth division almost drops the average to the ”acceptable” category of 1.50. These NWRs are shown graphically in Figure 20.

The conclusion is self evident. Adding another division makes a significant improvement to all forms of equity metrics.

C. Cap

In this section we analyze the presence or absence of a cap. Of the eight experiments, four have no cap in place and four have a cap of 1000 students. Figure 21 summarizes the presence/absence of a cap. With a cap, the number of schools with an acceptable NWR increases by 20 while the number

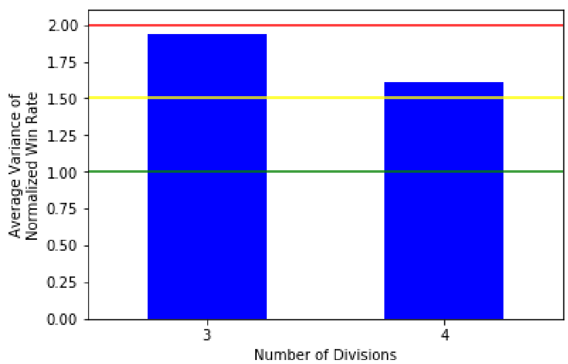

Fig. 20. 3 vs 4 Divisions: Normalized Win Rate Deviation

of unacceptable schools drops by 43. Similarly, the overall average normalized win rate decreases from 1.88 to 1.67 with the addition of a cap.

Cap Size	Num Accept	Number Marg	Number Unaccept	Normalized Win Rate
none	200.50	133.25	170.25	1.883
1000	233.75	161.50	127.75	1.667

Fig. 21. Average Results, No Cap vs Cap=1000

While the normalized win rate metrics *appear* to improve only moderately under caps, we must keep in mind that the presence of caps really only affects the schools in Division I. When we look at Division I, we see that the presence of caps dramatically changes the average NWR. The average normalized win rate falls from 3.673 to 1.898 when the current division process includes a cap, a 48% improvement in Div I.

Cap Size	Normalized Win Rate
none	3.673
1000	1.898

Fig. 22. Average Results **Div I only**: No Cap vs Cap=1000

There are other highly visible improvements with caps in place. For example, using the current division method (Equal Numbers with three divisions), Mason wins 850 state titles in simulation. When we introduce caps, Mason-A and Mason-B (the two subteams created when Mason splits) win 295 titles *combined*! Mason's normalized win rate falls from 5.75 when alone to an average of 1.9 for the two subteams. In general, introducing caps drops the Max NWR from 5.75 to 2.41 and raises the Min NWR from 0.026 to 0.126. Clearly, the cap concept greatly curtails the excessive win rates of the very largest programs and helps to raise the performance of the very smallest programs; caps greatly improve the competitive equity for schools in the "tails" of the distribution. This shows up prominently in Division I which suffers from the greatest competitive inequity.

D. Equal Numbers vs Equal Relative Difference

As we have found Equal Absolute Difference to be unworkable, that leaves two general methods: Equal Numbers versus Equal Relative Difference. There are four experiments of each method. Figure 23 shows the average equity metrics for these two groups of experiments. As can be seen, there is statistically no difference in the *average* of the two methods, they perform about equally well on average. However, looking more closely at the individual experiments (see Table of Figure 24), we do see that among the eight distinct experiments, there is a slight advantage for E12 over the others, which features the method of Equal Relative Difference. We also see in this table, that Equal Relative Difference outperforms Equal Numbers *only* when there is a cap in place. Otherwise, Equal Numbers fares better.

Method	Num Accept	Number Marg	Number Unaccept	Normalized Win Rate
EqNum	219.25	146.25	148.0	1.75975
EqRelDiff	215.00	148.50	150.0	1.79000

Fig. 23. Average Results: Equal Numbers vs Equal Relative Difference

VI. Conclusion

Our goal in this work is to explore alternatives to the default division process that could improve the competitive equity in Ohio high school girls' cross country. In our sensitivity analysis we discovered three important conclusions:

- Competitive equity is not affected by the absolute difference in enrollment numbers between different XC programs. As long as the enrollment ratio between the school's is fixed, it matters not how much larger one school is than another.
- Competitive equity is significantly anticorrelated to the relative difference in enrollment numbers between different XC programs. The enrollment ratio between schools has a significant impact on the ability of teams to compete fairly. Unfortunately, the inequity arises quickly, even with ratios of 1.25 or 1.5.
- There is a surprising effect of the *number* of teams in the competition. Adding more teams to a competitive event actually worsens the competitive equity. This suggests that in situations where there are large enrollment ratios, we should avoid large meets (more than five teams) if possible.

Unfortunately, the current OHSAA process includes both of these negative characteristics. The enrollment ratio between large and small schools in Division II and Division III is about 2.0; it is even worse in Division I with a ratio of 4.0. The process leading to a state championship title features a sequence of three larger meets (with 15 to 20 teams), thereby greatly amplifying the effects of enrollment disparity not just once, but three times in succession. Each of these three meets acts like a filter to cut out the smaller programs.

We then propose and explore a total of 12 models (including the baseline current model) that alter or add features to the division process. These 12 models feature three different methods of establishing break points for divisions:

- The Method of Equal Numbers: ensuring that each division has the same number of teams.
- The Method of Equal Absolute Difference: ensuring the enrollment difference between the largest and smallest teams in each division is the same.
- The Method of Equal Relative Difference: ensuring the enrollment ratio between the largest and smallest teams in each division is the same.

We discovered the Method of Equal Absolute Difference produces divisions that have wildly different numbers of teams (7 teams for Div I, 424 teams for Div III) and thus did not implement or test division processes with this method. The Method of Equal Relative Difference produces divisions of different sizes, but not so drastically as Equal Absolute Difference. We found a slight improvement in competitive equity using Equal Relative Difference over Equal Numbers, but only if a cap is included as well; otherwise, Equal Numbers performs slight better.

We also tested divisions processes with a cap. Teams from schools with more than 1000 female students would subdivide their runners into two teams (effectively halving the school enrollment population). The inclusion of a cap improved the competitive equity in all scenarios. The improvement was modest for an overall average NWR of 11%, but the improvement substantial within Division I, showing a 48% improvement in NWR.

Finally we tested a set of methods using four divisions instead of the baseline three divisions. Here again we saw an unambiguous improvement in all scenarios where we increase the number of divisions. The improvement in Normalized Win Rate deviation was more significant at 17%. We hypothesize, though did not test, that creating even more divisions would further improve competitive equity substantially.

A. Recommendations

From our test results, we propose the following recommendations for improving the division process in Ohio XC.

1) **Increase the Number of Divisions**: There is clear evidence that the competitive inequity in high school XC racing arises from significant enrollment variation of schools within divisions. There is simply no way to mitigate these effects. Increasing the number of divisions lowers the ratio between the largest and smallest schools in each division – the key factor in competitive equity.
2) **Manage the Largest Schools**: There are just a small handful of schools with large enrollments and yet they dominate the XC landscape, both in simulation and in reality. The top 20 largest programs comprise 3.9% of the schools yet they win over 51% of the state titles in simulation. Competitive equity would improve greatly if there were some way to manage these largest programs. A cap would be one method, though it is not likely to be politically favorable. A second way would be to create a separate division with just the largest group of schools – to isolate them from the other 95% of the schools which are much smaller. Isolating the largest programs will have a great effect on improving equity for the other schools.
3) **Retain Equal Numbers**: There is not enough of a significant advantage to the Method of Equal Relative Difference to overcome the administrative difficulties of managing divisions of different sizes. Equal Numbers would work significantly better with some combination of the two prior recommendations. The exception might be a separate smaller division that includes only the handful of very large schools (perhaps those with enrollments above 900), and then use Equal Numbers for the remaining schools in three other divisions.

ACKNOWLEDGMENTS

The authors would like to thank the Anderson Scholarship Fund and the William G. Bowen and Mary Ellen Bowen Endowed Fund for funding this work. Thanks to Granville Cross Country Coach Bart Smith and former coach Dave Agosta for providing data and insights as to Ohio's process of creating divisions.

REFERENCES

[1] Kretchmar, Matt. *The Effect of School Size on Cross Country Performance*. Submitted for publication. 2019.
[2] Aristotle, *Nicomachean Ethics*. New York, NY: Macmillan Publishing Company, 1962.
[3] FloTrack: http://www.wfltrack.org/cross/resources/HowToScoreXC.pdf
[4] Health Line: https://www.healthline.com/health/average-mile-time
[5] Liu, Jun S. *Monte Carlo Strategies in Scientific Computing*. New York, NY: Springer-Verlag. 2001.
[6] Ohio High School Athletic Association: https://ohsaa.org.
[7] OHSAA XC: https://www.ohsaa.org/sports/cc
[8] Ohio School Enrollment Data: http://education.ohio.gov/Topics/Data/Frequently-Requested-Data/Enrollment-Data
[9] Kovalchik, Stephanie Ann, and Martin Ingram. Estimating the duration of professional tennis matches for varying formats. *Journal of Quantitative Analysis in Sports*. Vol 14. No 1. 20018.
[10] Newton, Paul K., and Kamran Aslam. Monte Carlo Tennis: A Stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports*. Vol 5. No 3. 2009.
[11] Freeze, R. Allan. An Analysis of Baseball Batting Order by Monte Carlo Simulation. *Operations Research*. Vol 22. No. 4. Operations ResearchVol. 22, No. 4. 1974.
[12] Lewis, Michael M. *Moneyball: The Art of Winning an Unfair Game*. New York: W. W. Norton. 2003.
[13] Outdoor Industry: https://outdoorindustry.org/press-release/2015-strava-insights-show-cycling-and-running-landscape-in-the-u-s/.
[14] Runner's World: https://www.runnersworld.com/news/g25333911/strava-annual-report-running-statistics/.

Exp	Count of Schools			Normalized Win Rate		
	Acceptable	Marginal	Unacceptable	Variance	Max	Min
E1	162	137	205	2.073	5.747	0.026
E2	258	128	118	1.627	4.700	0.131
E3	182	165	176	1.808	2.401	0.126
E4	275	155	93	1.531	2.377	0.161
E5	na	na	na	na	na	na
E6	na	na	na	na	na	na
E7	na	na	na	na	na	na
E8	na	na	na	na	na	na
E9	182	114	208	2.077	3.503	0.032
E10	200	154	150	1.756	2.381	0.110
E11	195	162	166	1.803	2.410	0.055
E12	283	164	76	1.524	2.111	0.238

Fig. 24. Table of All Experimental Results

VII. Appendix I: Graphs of Experimental Results

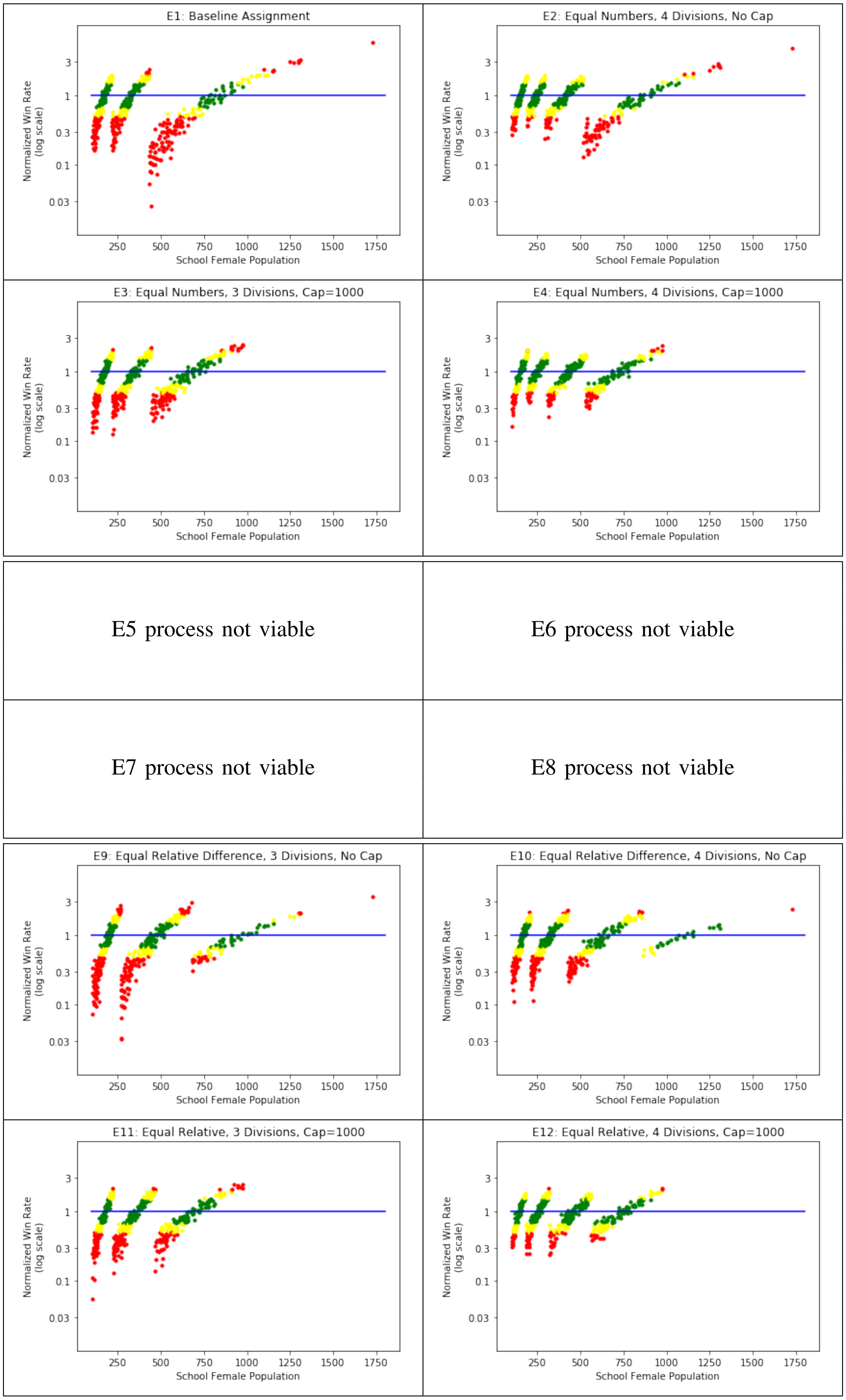

Fig. 25. All Graph Results