

# Comparison of Data Screening Methods for Evaluating School-Level Fitness Patterns in Youth: Findings from the NFL PLAY 60 FITNESSGRAM Partnership Project

Pedro F. Saint-Maurice<sup>1,2\*</sup>, Gregory J. Welk<sup>1</sup>, Yang Bai<sup>1</sup>, Kelly Allums-Featherston<sup>3</sup>

<sup>1</sup>Department of Kinesiology, Iowa State University, Ames, USA

<sup>2</sup>School of Psychology-CIPsi, University of Minho, Braga, Portugal

<sup>3</sup>NFL PLAY 60 FITNESSGRAM® Project, The Cooper Institute®, Dallas, USA

Email: [pedrosm@iastate.edu](mailto:pedrosm@iastate.edu), [gwelk@iastate.edu](mailto:gwelk@iastate.edu), [ybai@iastate.edu](mailto:ybai@iastate.edu), [kallumsfeatherston@cooperinst.org](mailto:kallumsfeatherston@cooperinst.org)

Received 9 September 2014; revised 22 October 2014; accepted 20 November 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

**Background:** There has been a great interest in tracking health-related fitness across the United States. The NFL PLAY 60 FITNESSGRAM Partnership Project (NFL P60FGPP) is a large participatory research network that involves the surveillance of fitness among more than 1000 schools spread throughout the country. Fitness data are collected by school staff and therefore these data can vary in quality and representativeness. Therefore, careful screening procedures are needed to ensure that the data can reflect actual patterns in the schools. This study examined the impact of different data screening procedures on outcomes of aerobic fitness (AF) collected from the NFL P60FGPP. **Methods:** Data were compiled from 149,101 youth from 504 schools and were processed using the established age- and gender-specific AF FITNESSGRAM health-related standards. Data were subjected to three different screening procedures (based on grade size and boy-to-girl ratio per grade). Linear models were computed to obtain unadjusted and adjusted (for age, BMI-Z, and socio-economic status) estimates of % youth in the Healthy Fitness Zone (HFZ) in order to determine if, 1) there were differences in % in the HFZ and 2) if differences could be explained by changes in the representativeness of the sample due to the different data screening procedures. **Results:** Depending on the screening procedure used, the final sample ranged from 96,999 (no screening) to 46,572 youth (most stringent criteria). The proportion of youth achieving appropriate levels of AF ranged from 56% to 61% with unscreened data resulting in consistently lower percentages of youth achieving the standard ( $P < 0.05$ ). Overall, these differences were not explained by possible changes in demographic characteristics as the result of applying different screening criteria. **Conclusions:** The findings demonstrate the importance of establishing appro-

\*Corresponding author.

**How to cite this paper:** Saint-Maurice, P.F., Welk, G.J., Bai, Y. and Allums-Featherston, K. (2014) Comparison of Data Screening Methods for Evaluating School-Level Fitness Patterns in Youth: Findings from the NFL PLAY 60 FITNESSGRAM Partnership Project. *Open Journal of Preventive Medicine*, 4, 876-886. <http://dx.doi.org/10.4236/ojpm.2014.411099>

appropriate screening procedures that maximize sample size while also ensuring generalizability of the findings.

## Keywords

**Aerobic Fitness, Assessment, Data Reduction, Public Health, Surveillance**

## 1. Introduction

A recent study of 27 countries has documented global declines in aerobic fitness performance ( $-0.36\%$  per year) over the past 50 years [1]. This study highlighted issues with public health surveillance and substantiated the importance of youth physical activity and fitness promotion on an international level.

Schools have been increasingly emphasized as a promising target for coordinated programming [2] [3] and studies have evaluated numerous school-based interventions [4] [5] and policies focused on youth fitness [3] [6] [7]. Fitness testing has been a mainstay of Physical Education (PE) programs in the United States for over 50 years [8], but existent fitness surveys [9] [10] are out of date due to the secular changes of youth fitness [1]. Interestingly, many states and large districts now mandate, fund, and/or promote the systematic collection of health-related fitness data for youth fitness surveillance (e.g., California, Texas, Georgia, Delaware, New York City).

The FITNESSGRAM<sup>®</sup> battery is one of the most commonly used fitness batteries across the globe [11] [12]. The FITNESSGRAM program helped to shift the focus of fitness testing from performance-related to health-related fitness and from norm-referenced to criterion-referenced evaluation standards [11]. In addition to personalized reporting, the new web-based platform makes it possible to compile and track youth fitness data by class, school, district and state level, along with printed personalized assessment reports for individuals. A series of recent studies published in the American Journal of Preventive Medicine refined the criterion-referenced standards for aerobic fitness and body composition using nationally-representative data from the National Health and Nutrition Examination Survey. The revised standards for aerobic fitness [13] and body composition [14] have documented utility for detecting risks of metabolic syndrome in youth and have been shown to be related with the fulfillment of physical activity guidelines [15]. This allows the school-based FITNESSGRAM assessment to provide valuable information about levels of health-related fitness in youth.

One example of a large-scale application of FITNESSGRAM was the Texas Youth Fitness Study [16]. Results from the ongoing FITNESSGRAM adoption in Texas have been published in a series of studies [13] [17]-[20]. The supplement included detailed reports of fitness results [17] as well as a controlled study that demonstrated that trained teachers could provide valid and reliable data on youth fitness [18]. These results demonstrate the potential for trained teachers to adopt FITNESSGRAM testing for public health surveillance. In addition, the Presidential Youth Fitness Program ([www.presidentialyouthfitnessprogram.org](http://www.presidentialyouthfitnessprogram.org)) has now established FITNESSGRAM program as the exclusive national youth fitness battery.

The widespread adoption of FITNESSGRAM (both in the US and internationally) opens up exciting opportunities to systematically study age and gender patterns of youth fitness with the same test. However, there are many complex issues that must be considered when using field-based tests for tracking fitness levels at a large-scale (e.g., district, state, national). Data collected from schools can vary in quality and representativeness. Therefore, careful data screening procedures are needed to ensure that the data can reflect actual patterns in the schools. Some of these concerns have been studied in other health surveillance research areas [21]-[24].

The purpose of this study was to systematically evaluate different data screening procedures on school level estimates of fitness outcomes collected from local schools spread throughout the US. The impact of different data screening procedures was examined in a sample of over 500 schools from 22 different states.

## 2. Methods

### 2.1. Design and Sample

Institutional Review Board approval was granted from the Cooper institute and Iowa State University. Data for the present study were obtained through a participatory research project called the NFL PLAY 60 FITNESS-

GRAM Partnership Project (NFL P60FGPP). The NFL P60FGPP, launched in 2009, provides training and support to a network of over 1100 schools/sites (35 site licenses per each of the 32 NFL franchises) [25]. Schools are provided training materials that include the FITNESSGRAM protocol and philosophy and are encouraged to use various NFL PLAY 60 resources to promote physical activity and healthy eating in students. FITNESSGRAM records from the participating schools are compiled through web-based servers and these data are tracked over time to evaluate fitness patterns and the impact of school programming under real world conditions.

The data for the present study were collected in the Spring of 2012 and extracted from the project database in the Fall of 2012. The dataset included 149,101 student records from 504 schools from 32 different NFL franchise cities) in 22 states and one region (New England) in the US. There were 7546 cases with missing demographic information (age, gender, grade scores and 21,940 that were either missing or had unfeasible aerobic fitness scores so these cases were excluded (individual-level screening). An additional 22,616 cases were removed because they were out of the targeted age range (10 - 18) for aerobic fitness evaluation. This resulted in a final sample of 96,099 records that had complete and clean data on the outcome of interest.

## 2.2. The FITNESSGRAM® Assessment

The established FITNESSGRAM battery includes a variety of practical, field-based assessments for each of the key dimensions of health-related fitness (aerobic capacity, body composition, and musculoskeletal fitness and flexibility) but the focus in the present study was on aerobic fitness measures. The FITNESSGRAM battery provides schools with three different assessments of aerobic fitness, a progressive 20 m shuttle run (PACER), the one-mile run test (MRT), and a 15-meter PACER test (a modified version of the PACER test). These assessment's scores are equated to PACER laps [26]-[28] and then predicted maximal  $\text{VO}_2$  is evaluated using the established health-related standards. Youth that achieve the standard are placed into the Healthy Fitness Zone® (HFZ) while youth falling below this value are placed in the Needs Improvement Zone (NIZ). Schools that utilize the FITNESSGRAM web-based software have the ability to generate personalized fitness reports and aggregate level reports that schools can use to view the percentage of youth achieving the HFZ. However, the focus in the present study was on decisions that would influence interpretations of compiled school level data for public health surveillance.

## 2.3. Data Processing

Data analyses were conducted in the Fall of 2013. Fitness data from the NFL PLAY 60 campaign are hierarchically structured with individuals nested in grades within a particular school and further nested by franchise.

De-identified fitness data files were first exported from the web servers and cleaned using standard procedures to ensure the quality of the individual records. Participants were excluded if they were missing demographic information such as age, gender, or grade information, or did not have aerobic capacity scores. Participants were also excluded if they had abnormal values for aerobic capacity (e.g., values = 0, or out of range scores). The cleaned data were then aggregated by gender and grade and then screened using three different approaches that varied in rigor:

- Screening A (conservative): This screening protocol was the most conservative. Cases were removed if there was an unbalance between the number of boys and girls (a gender ratio greater than 1.2 or 12:10 ratio) and if there were less than 60 students per grade.
- Screening B (intermediate): This screening protocol was defined to be less conservative than protocol A. The ratio boy:girl criteria was the same but cases were removed if there were less than 30 students per grade.
- Screening C (liberal): This was the most liberal protocol. Cases were only removed if there was an excessive boy:girl unbalance (a gender ratio greater than 2.0 or 20:10 ratio) and if there was an excessively small sample of students per grade (grades with less than 15 students).

The four different (screened and unfiltered) data sets were then processed to compare the impact on school level fitness outcomes. We used the percentage of students per grade meeting the Healthy Fitness Zone as our outcome variable since this is a popular indicator used in youth fitness research. This was calculated separately for boys and girls using the formula below:

$$\left( \frac{\text{Number of students meeting the HFZ for AC at each grade per school}}{\text{Total number of students at each grade per school}} \times 100 \right).$$

## 2.4. Data Analysis

As part of our analysis, we used aggregated grade-level data to provide a visual illustration (histograms) of how the raw data (no filters) was distributed for the ratio boy/girl per grade and total participants per grade, screening variables. The impact of each screening decision on the initial sample size and outcome scores was determined visually using histograms along with skewness and kurtosis values as indicators of the shape of the distribution.

We were particularly interested in the effect of data screening on state-level outcomes. Therefore, the effect of screening protocol was determined on aggregated state-level data using a within-subjects design with “Percent Meeting the Healthy Fitness Zone” (% Meeting the HFZ) as the outcome variable. We computed two linear models: The first included only one predictor (screening protocol) and examined if there were statistically significant differences between each screening protocol. The second linear model included average state-level age, BMI-Z scores, and socio-economic status (SES) indicators and respective interactions with screening protocol, to determine if differences between screening protocols (model 1) could be explained by changes in demographic characteristics in each screening protocol sample. Differences in the output obtained from the two models would indicate that data screening protocols can result in exclusion of important segments of the population being studied which might lead to important fluctuations on state-level estimates of health-related fitness. Socio-economic status was calculated as the percentage of participants at each school eligible for free or reduced lunch and the three predictor variables were centered at the sample median score. The solution for fixed effects resultant from each model above was followed by pre-determined contrasts between the raw outcome scores (no filters) and each of the screening protocols. Differences in least square means were tested using 95% confidence intervals ( $P < 0.05$ ).

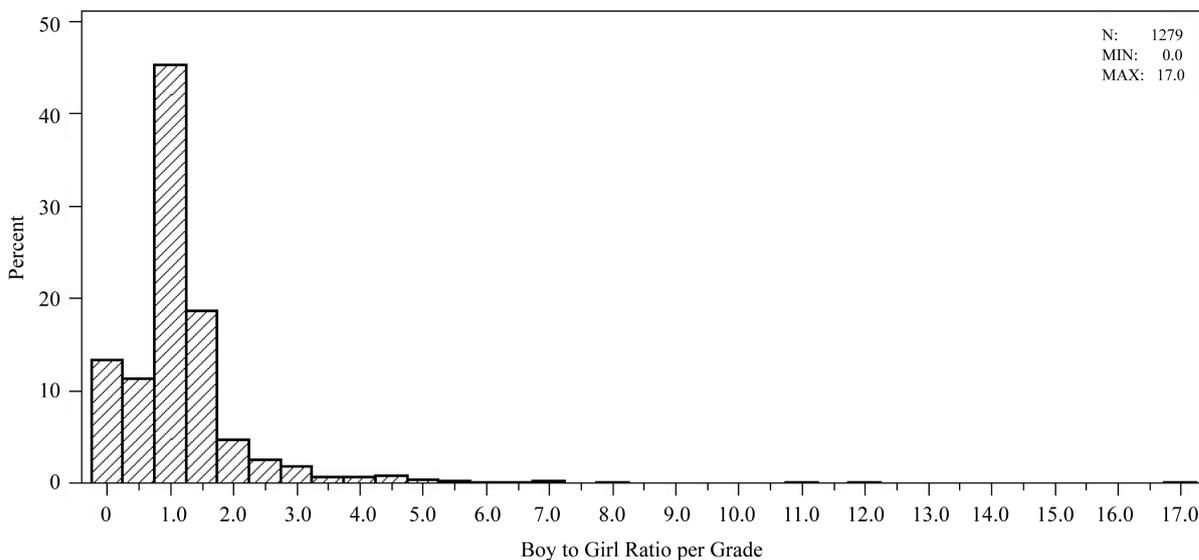
## 3. Results

The number of excluded grades varied across the three screening protocols depending on the stringency of the criteria (see **Table 1**). When no filters were applied, the final sample resulted in 1279 grades. The boy per girl ratio for each school grade ranged from 0 (*i.e.*, indicating some school grades just had either boys or girls with valid CVF scores) to 17 (*i.e.*, indicating a ratio of 17 boys per girl per school grade with valid CVF scores) (**Figure 1(a)**). The total number of students per school grade ranged from 1 to 867 with approximately 30% of the school grades having less than 15 students per grade (**Figure 1(b)**).

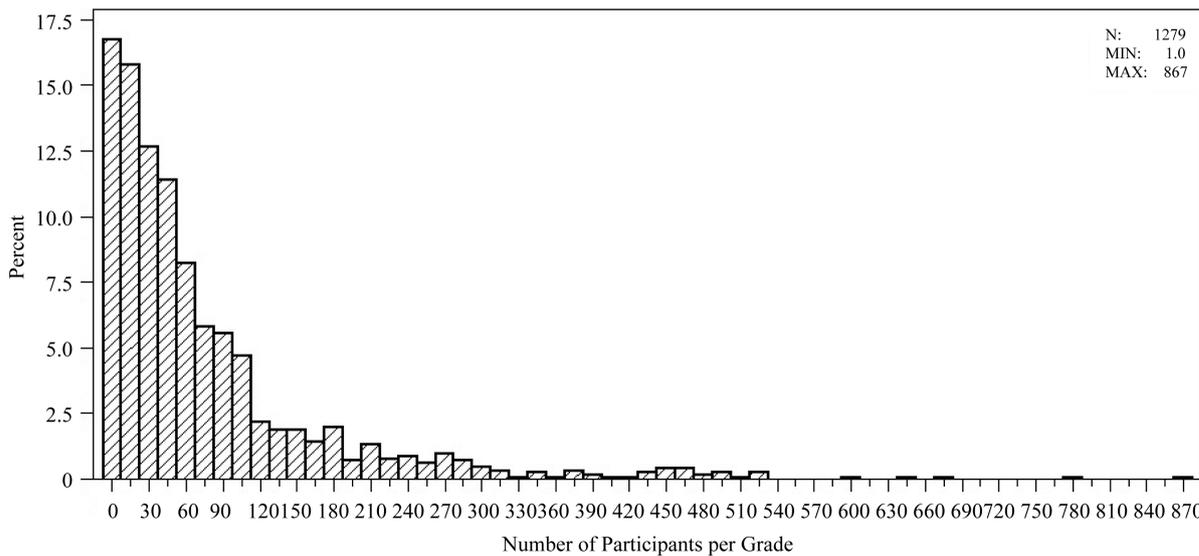
**Table 1.** Flow of sample size by screening protocol.

Screening	Method			
	No filter	Screening A	Screening B	Screening C
N individual		149,101 (100%)		
N grades		1804 (100%)		
N schools		504 (100%)		
N franchises		32 (100%)		
<b><u>IL screening</u></b>		96,999 (65.1%)		
<b><u>GL screening</u></b>	NA	1279 (70.9%)	1279 (70.9%)	1279 (71.2%)
Ratio boys/girls	NA	482 (26.7%)	482 (26.7%)	981 (54.4%)
Students per grade	NA	272 (15.1%)	388 (21.5%)	844 (46.8%)
<b>Final sample</b>				
Individuals	96,999 (65.1%)	46,572 (31.2%)	53,173 (35.7%)	88,712 (59.5%)
Grades	1279 (70.9%)	272 (15.1%)	388 (21.5%)	844 (46.8%)
Schools	404 (80.2%)	168 (33.3%)	234 (46.4%)	349 (69.2%)
Franchises	31 (96.9%)	31 (96.9%)	31 (96.9%)	31 (96.9%)

IL: individual level screening (phase I); GL: grade level screening (phase II); NA: not applicable; screening A: boy per girl ratio = 1.2 and total number of participants per grade = 60; screening B: boy per girl ratio = 1.2 and total number of participants per grade = 30; screening C: boy per girl ratio = 2.0 and total number of participants per grade = 15.



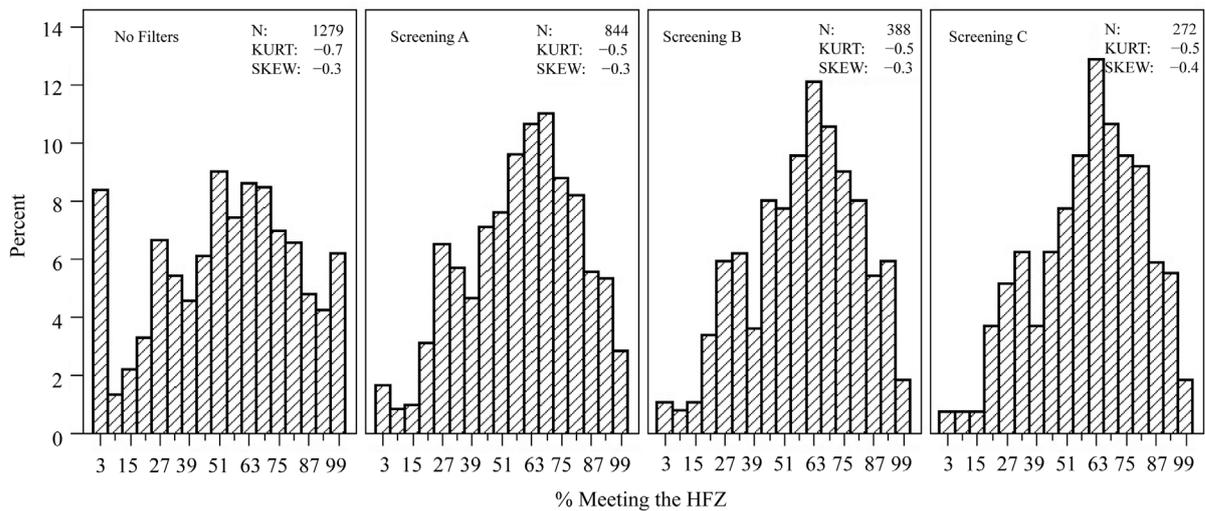
(a)



(b)

**Figure 1.** (a) Distribution of boy to girl ratio per grade per school in the raw (unfiltered) data (n = 1279 grades). The minimum (MIN) value for this distribution was equal to 0 (indicating some grades only had either boys or girls) while the maximum (MAX) was 17 (indicating some grades had a 17 boy to girl ratio); (b) Distribution of total number of participants per grade per school in the raw (unfiltered) data (n = 1279). The minimum (MIN) value for this distribution was 1 (indicating some grades only had 1 participant) while some grades had 867 participants (MAX).

The impact of different screening protocols on grade-level aerobic fitness estimates was first examined based on standard indicators of sample distribution. **Figure 2** indicates that the more stringent the screening protocol was the more normally distributed the indicators of fitness were. When no filters were applied, there was a higher prevalence of extreme scores. For example, the “no filters” histogram indicated that approximately 8% of the total number of grades had 0% to 3% of the students meeting the HFZ while 6% of the total grades had 99% of their students meeting the HFZ for aerobic fitness. The prevalence of scores at the two ends of the spectrum dropped to approximately 2% when filters were applied. Values for kurtosis were lower for the unscreened sample and similar among the three screening procedures (*i.e.*, indicating that scores were more spread out when no filters were applied). There were no differences in the skewness between the protocols except that screening protocol C had slightly lower skewness values (**Figure 2**).



**Figure 2.** Distribution of the percent of participants meeting the Healthy Fitness Zone for the raw data (left: no filters) and by screening protocol. Values for kurtosis (KURT) and skewness (SKEW) are provided in the top right side of each distribution.

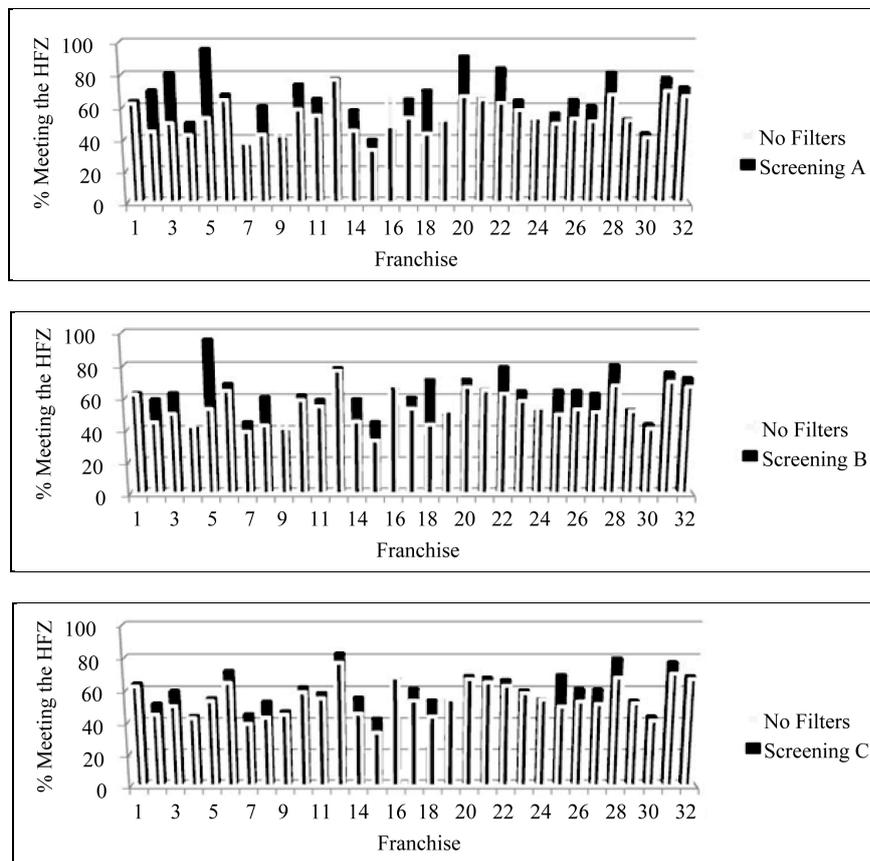
Grade-level indicators of fitness were aggregated by franchise ( $n = 31$ ) to examine the effect of data reduction decisions when reporting nationwide fitness results. **Figure 3** illustrates how the average estimates of students meeting the HFZ resultant from each screening protocol (A, B, and C) changes when all the raw data were included for analysis. Then visual patterns suggested that including all the raw data when processing fitness data would lead to lower estimates of aerobic fitness in most of the states. The black bars in **Figure 3** are more visible as the protocol becomes more stringent. If no filters were to be used, the proportion of students meeting the HFZ in most of the states would be less or equal to 60%. This value fluctuates when some screening criteria are used and several states actually reach the 80% mark when the most stringent protocol was used (**Figure 3**).

The solution for the fixed effects model indicated a statistically significant and linear effect of protocol [ $F(3, 30) = 6.31, P < 0.01$ ]. The more rigorous the protocol was the greater the discrepancy with raw (unscreened) data. The unadjusted state-level proportions of students meeting the HFZ were 55.3% (1.9%), 58.5% (1.9%), 61.0% (2.3%), and 62.4% (2.8%), for raw, protocol C, protocol B, and protocol A, respectively. Pairwise comparisons of each protocol with the raw data revealed statistically significant differences and wide confidence intervals for the discrepancy in estimates of aerobic fitness ( $[t_{\text{Protocol C}}(30) = -3.86, 95\% \text{ CI: } -4.8\%, -1.5\%; P < 0.01; t_{\text{Protocol B}}(30) = -3.17, 95\% \text{ CI: } -9.2\%, -2.0\%; P < 0.01; t_{\text{Protocol A}}(30) = -3.12, 95\% \text{ CI: } -11.7\%, -2.4\%; P < 0.01]$ ). The same comparisons, when adjusted for average state-level age, SES, and BMI-Z scores, showed a similar pattern for the effect of protocol [ $F(3, 30) = 3.25, P = 0.04$ ]. The average proportion of youth meeting the HFZ was equal to 56.1% (1.6%), 58.4% (1.8%), 59.2% (2.3%), and 60.9% (2.4%), for raw, and protocols C, B, and A, respectively. Pairwise comparisons deemed statistically significant ( $[t_{\text{Protocol C}}(30) = -2.65, 95\% \text{ CI: } -4.1\%, -0.5\%; P = 0.01; t_{\text{Protocol A}}(30) = -2.56, 95\% \text{ CI: } -8.5\%, -1.0\%; P = 0.02]$ ) except for protocol B [ $t(30) = -1.86, 95\% \text{ CI: } -6.5\%, 0.3\%; P = 0.07$ ] (**Figure 4**).

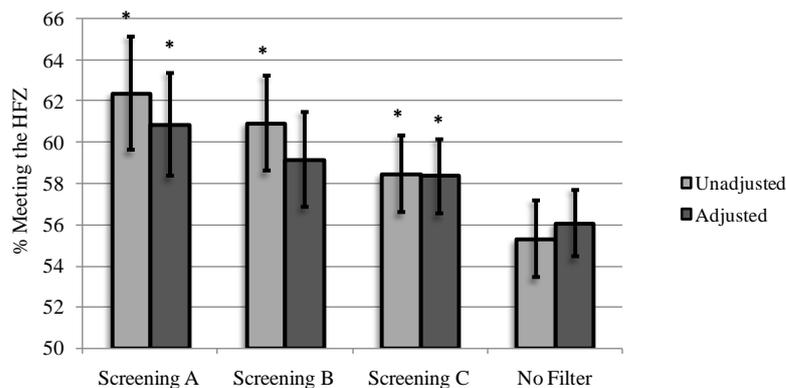
#### 4. Discussion

There are a number of large-scale fitness surveillance initiatives taking place in states and nations across the globe; however, little consideration has been given to the techniques used to process and control the quality of the data. As shown in this study, the distribution of fitness scores when data were not filtered can result in overall lower estimates of youth meeting the HFZ. We characterized the distribution of field-based fitness assessment scores in a large national cohort of schools involved in the NFL P60FGPP, but the conclusions and implications would have relevant impact for district, state or national evaluations.

We focused the analyses on aerobic fitness because it is widely considered to be the most important component of health-related fitness and is almost universally used in youth fitness batteries. The FITNESSGRAM battery recommends the use of the PACER test and this assessment is based on the original 20 m shuttle run test



**Figure 3.** Pairwise comparisons of the average US state (franchise) percent of students meeting the Healthy Fitness Zone. The raw data average % in HFZ was set as the reference (white bars) while each screening protocol was defined using black bars.



**Figure 4.** Adjusted and unadjusted average number of participants meeting the Healthy Fitness Zone for the raw data (unfiltered) and each screening protocol. Statistically significant differences are indicated with \* and represent pairwise comparisons (using adjusted or unadjusted least square means) between the raw data and each screening protocol.

that is widely used in European test batteries (Eurofit) and other national batteries [29]. An advantage of the PACER test is that it essentially replicates the timing and structure of standard lab-based maximal fitness tests [30]. It was recommended by the Institute of Medicine as the primary field-based assessment of aerobic fitness for youth and has good psychometric and motivational properties for school-based assessments [12] [31].

The PACER test, like other assessments in the FITNESSGRAM battery, was developed primarily for fitness education and youth fitness promotion. Youth typically receive personalized reports about their level of fitness and feedback on how to maintain or improve it. Educating students about their level of health-related fitness is the primary function of youth fitness testing but the FITNESSGRAM advisory board has also endorsed the use of “*institutional testing*” as appropriate uses of fitness data [32]. Districts often track data to examine the impact of curricular changes or the long term impact of programming on youth fitness. States may be more interested in evaluating the impact of school environments or policies on youth fitness outcomes. For example, an evaluation of the Texas Youth Fitness study examined whether various school level factors (*i.e.*, school physical education policy, school resources, physical education duration and frequency, teachers’ training and testing experience) could explain the variability in fitness results observed across the state [20]. In either example, the most important consideration is for standardization in the processes used for screening and processing the data. Based on the present analyses, the simple inclusion of all available data would likely lead to spurious findings when used to explain age and gender patterns of fitness.

There are no definitive guidelines to determine what the “correct” screening criteria would be. The use of more stringent criteria would restrict the available sample and possibly increase the internal validity of the findings. However, the restrictions could jeopardize the external validity of the findings since it would lead to a less-representative sample population. The complexities of these issues are compounded when trying to understand differences across schools since there is considerable variability in the nature and size of schools. A sample of 15 children per grade may be a small number of students in a large school but it could be the full grade contingent in a small rural school. Therefore, it may be important to also consider the percent of the available students tested rather than simply the number of students. This discussion has important implications for future surveillance reports on physical fitness. It is important to facilitate and promote the integration of fitness assessment in schools across the country in order to improve the quality of the data for surveillance. These trends can provide important information on the state of art of surveillance of a specific health indicator [33].

Additionally, it is important to consider the boy-to-girl ratio when using aggregated data as representative of a grade. It is well known that both absolute and relative aerobic capacity is higher in boys, particularly, during late adolescence [34]. The FITNESSGRAM criterion-referenced standards account for these differences; however, gender differences can also be expected in the proportion of youth meeting the recommended levels of aerobic fitness [17]. These findings support our decision to include a gender ratio screening criteria. Any unbalance at a specific grade can lead to flawed comparisons when using aggregated data. At this point we were not able to determine which gender ratio would be the most appropriate but our study shows that the inclusion of this requirement has some implications for aerobic fitness outcomes.

A strength of the present study is the large sample of schools, this made it possible to simulate issues that would arise when aggregating data at both the state and national level. Limitations of the analyses are the use of only one fitness measure and the focus on only 3 key screening methods. The screening criteria methods used in this study are more closely related to the definition of coverage (*e.g.*, extent to what students assessed are representative of a particular school or state). Coverage can be seen as an indicator of the quality of the data [22] and can be improved when several of the study design decisions (*e.g.*, random selection, stratified sample selection) are within the control of the researchers [35]-[37]. It is possible that more robust analyses could optimize criteria for specific decisions and include other dimensions of data quality. If consensus can be reached on appropriate screening procedures for “*naturalistic*” studies designs (*i.e.*, non-random sample selection) it could enable more effective comparisons across districts, states and nations. The broad adoption of FITNESSGRAM across the United States and many other countries offer considerable promise for advancing understanding of youth fitness outcomes. However, the results of the present study demonstrate the importance of selecting an appropriate screening protocol to ensure appropriate interpretations.

## 5. Conclusion

The present study used a large national sample to demonstrate the impact of the quality of the data on state-level estimates of health-related fitness. The NFL PLAY 60 FITNESSGRAM Partnership has been successfully in implementing a sustainable strategy for the surveillance of fitness across the country. This initiative will allow states to have a better understanding of youth fitness levels and their implications for public health. However, this approach relies on local school efforts to assess children and record their fitness data using appropriate pro-

cedures which might lead to a lack of standardization of data collection/record procedures. It can also lead to poorly represented sites/schools/states if only a small set of students are assessed without any additional information on the selection criteria used to determine this sample (selection bias). This will most likely affect the quality of the data. It is challenging to define or quantify the quality of large scale data but our study provides evidence that the quality of the data, namely, coverage, can vary between schools and that unscreened fitness data can result in a higher prevalence of unsound estimates of aerobic fitness. We demonstrated that unscreened data led to lower levels of aerobic fitness in youth when compared to the same data when some screening procedures were used. Importantly, our results also demonstrated that more stringent screening procedures can reduce the available sample to a great degree; however, it did not affect the representativeness of the sample being considered. The NFL PLAY 60 FITNESSGRAM initiative provides the most up to date information about youth fitness levels across the country. This initiative provides a unique opportunity to explore some of the challenges associated with large scale tracking of health-related indicators. We suggest that screening procedures be considered when quality control is not sustainable or cannot be assured. This is a very common case in surveillance research.

## Acknowledgements

This work and the NFL PLAY 60 FITNESSGRAM Partnership Project were supported by a grant from the NFL Foundation (formerly NFL Charities) to provide 1120 FITNESSGRAM licenses, access and support for the web-based software, staff development, and support to a project evaluation team.

## Disclosures or Conflicts of Interest

No financial disclosures or conflicts of interest were reported by the authors of this paper.

## References

- [1] Tomkinson, G.R. and Olds, T. (2007) *Secular Changes in Pediatric Aerobic Fitness Test Performance: The Global Picture*. Karger Publishers, Berlin.
- [2] Wechsler, H., Devereaux, R.S., Davis, M. and Collins, J. (2000) Using the School Environment to Promote Physical Activity and Healthy Eating. *Preventive Medicine*, **31**, S121-S137.  
[http://ac.els-cdn.com/S0091743500906492/1-s2.0-S0091743500906492-main.pdf?\\_tid=36ef7636-607e-11e3-99a6-0000aab0f6b&acdnat=1386558275\\_169dcda2077b0b3ea0aa901d82ddea8a](http://ac.els-cdn.com/S0091743500906492/1-s2.0-S0091743500906492-main.pdf?_tid=36ef7636-607e-11e3-99a6-0000aab0f6b&acdnat=1386558275_169dcda2077b0b3ea0aa901d82ddea8a)  
<http://dx.doi.org/10.1006/pmed.2000.0649>
- [3] Dobbins, M., DeCorby, K., Robeson, P., Husson, H. and Tirilis, D. (2009) School-Based Physical Activity Programs for Promoting Physical Activity and Fitness in Children and Adolescents Aged 6 - 18. *The Cochrane Database of Systematic Reviews*, **2009**, Article ID: CD007651.
- [4] Christiansen, L.B., Toftager, M., Boyle, E., Kristensen, P.L. and Troelsen, J. (2013) Effect of a School Environment Intervention on Adolescent Adiposity and Physical Fitness. *Scandinavian Journal of Medicine & Science in Sports*, **23**, e381-e389. <http://dx.doi.org/10.1111/sms.12088>
- [5] Dobbins, M., Husson, H., DeCorby, K. and LaRocca, R.L. (2013) School-Based Physical Activity Programs for Promoting Physical Activity and Fitness in Children and Adolescents Aged 6 to 18. *The Cochrane Database of Systematic Reviews*, **2013**, Article ID: CD007651. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007651.pub2/abstract>
- [6] London, R.A. and Gurantz, O. (2013) Afterschool Program Participation, Youth Physical Fitness, and Overweight. *American Journal of Preventive Medicine*, **44**, S200-S207.  
[http://ac.els-cdn.com/S0749379712008689/1-s2.0-S0749379712008689-main.pdf?\\_tid=e0ab7dd4-607c-11e3-ab34-0000aacb361&acdnat=1386557701\\_8574e9adad682089c8ec3f7c9c059c86](http://ac.els-cdn.com/S0749379712008689/1-s2.0-S0749379712008689-main.pdf?_tid=e0ab7dd4-607c-11e3-ab34-0000aacb361&acdnat=1386557701_8574e9adad682089c8ec3f7c9c059c86)  
<http://dx.doi.org/10.1016/j.amepre.2012.11.009>
- [7] Beets, M.W., Beighle, A., Erwin, H.E. and Huberty, J.L. (2009) After-School Program Impact on Physical Activity and Fitness: A Meta-Analysis. *American Journal of Preventive Medicine*, **36**, 527-537.  
[http://ac.els-cdn.com/S0749379709001470/1-s2.0-S0749379709001470-main.pdf?\\_tid=3bdbcaec-607d-11e3-9215-0000aacb35f&acdnat=1386557854\\_f6eb8ec34aa8b9a07975639a914a028f](http://ac.els-cdn.com/S0749379709001470/1-s2.0-S0749379709001470-main.pdf?_tid=3bdbcaec-607d-11e3-9215-0000aacb35f&acdnat=1386557854_f6eb8ec34aa8b9a07975639a914a028f)  
<http://dx.doi.org/10.1016/j.amepre.2009.01.033>
- [8] Morrow Jr., J.R., Zhu, W., Franks, B.D., Meredith, M.D. and Spain, C. (2009) 1958-2008: 50 Years of Youth Fitness Tests in the United States. *Research Quarterly for Exercise and Sport*, **80**, 1-11.  
<http://www.ncbi.nlm.nih.gov/pubmed/19408462>

- [9] McGinnis, J.M. (1985) National Children and Youth Fitness Study. *Journal of Physical Education, Recreation and Dance*, **56**, 44-90. <http://dx.doi.org/10.1080/07303084.1985.10603682>
- [10] Ross, J.G. and Pate, R.R. (1987) The National Children and Youth Fitness Study II. *Journal of Physical Education, Recreation and Dance*, **58**, 50-96.
- [11] Plowman, S.A., Sterling, C.L., Corbin, C.B., Meredith, M.D., Welk, G.J. and Morrow, J.R. (2006) The History of FITNESSGRAM. *Journal of Physical Activity & Health*, **3**, S5-S20.
- [12] Castro-Pinero, J., Artero, E.G., Espana-Romero, V., Ortega, F.B., Sjostrom, M., Suni, J. and Ruiz, J.R. (2010) Criterion-Related Validity of Field-Based Fitness Tests in Youth: A Systematic Review. *British Journal of Sports Medicine*, **44**, 934-943. <http://www.ncbi.nlm.nih.gov/pubmed/19364756>  
<http://dx.doi.org/10.1136/bjism.2009.058321>
- [13] Welk, G.J., Laurson, K.R., Eisenmann, J.C. and Cureton, K.J. (2011) Development of Youth Aerobic-Capacity Standards Using Receiver Operating Characteristic Curves. *American Journal of Preventive Medicine*, **41**, S111-S116. <http://www.ncbi.nlm.nih.gov/pubmed/21961610>  
<http://dx.doi.org/10.1016/j.amepre.2011.07.007>
- [14] Laurson, K.R., Eisenmann, J.C. and Welk, G.J. (2011) Development of Youth Percent Body Fat Standards Using Receiver Operating Characteristic Curves. *American Journal of Preventive Medicine*, **41**, S93-S99. <http://www.ncbi.nlm.nih.gov/pubmed/21961618>  
<http://dx.doi.org/10.1016/j.amepre.2011.07.003>
- [15] Morrow Jr., J.R., Tucker, J.S., Jackson, A.W., Martin, S.B., Greenleaf, C.A. and Petrie, T.A. (2013) Meeting Physical Activity Guidelines and Health-Related Fitness in Youth. *American Journal of Preventive Medicine*, **44**, 439-444. <http://www.ncbi.nlm.nih.gov/pubmed/23597805>  
<http://dx.doi.org/10.1016/j.amepre.2013.01.008>
- [16] Morrow Jr., J.R., Martin, S.B., Welk, G.J., Zhu, W. and Meredith, M.D. (2010) Overview of the Texas Youth Fitness Study. *Research Quarterly for Exercise and Sport*, **81**, S1-S5. <http://www.ncbi.nlm.nih.gov/pubmed/21049832>  
<http://dx.doi.org/10.1080/02701367.2010.10599688>
- [17] Welk, G.J., Meredith, M.D., Ihmels, M. and Seeger, C. (2010) Distribution of Health-Related Physical Fitness in Texas Youth: A Demographic and Geographic Analysis. *Research Quarterly for Exercise and Sport*, **81**, S6-S15. <http://www.ncbi.nlm.nih.gov/pubmed/21049833>  
<http://dx.doi.org/10.1080/02701367.2010.10599689>
- [18] Morrow Jr., J.R., Martin, S.B. and Jackson, A.W. (2010) Reliability and Validity of the FITNESSGRAM: Quality of Teacher-Collected Health-Related Fitness Surveillance Data. *Research Quarterly for Exercise and Sport*, **81**, S24-S30. <http://www.ncbi.nlm.nih.gov/pubmed/21049835>  
<http://dx.doi.org/10.1080/02701367.2010.10599691>
- [19] Zhu, W., Welk, G.J., Meredith, M.D. and Boiarskaia, E.A. (2010) A Survey of Physical Education Programs and Policies in Texas Schools. *Research Quarterly for Exercise and Sport*, **81**, S42-S52. <http://www.ncbi.nlm.nih.gov/pubmed/21049837>  
<http://dx.doi.org/10.1080/02701367.2010.10599693>
- [20] Zhu, W., Boiarskaia, E.A., Welk, G.J. and Meredith, M.D. (2010) Physical Education and School Contextual Factors Relating to Students' Achievement and Cross-Grade Differences in Aerobic Fitness and Obesity. *Research Quarterly for Exercise and Sport*, **81**, S53-S64. <http://www.ncbi.nlm.nih.gov/pubmed/21049838>  
<http://dx.doi.org/10.1080/02701367.2010.10599694>
- [21] Ng, M., Gakidou, E., Murray, C.J. and Lim, S.S. (2013) A Comparison of Missing Data Procedures for Addressing Selection Bias in HIV Sentinel Surveillance Data. *Population Health Metrics*, **11**, 12. <http://www.ncbi.nlm.nih.gov/pubmed/23883362>  
<http://dx.doi.org/10.1186/1478-7954-11-12>
- [22] Walker, N., Garcia-Calleja, J.M., Heaton, L., Asamoah-Odei, E., Pomeroy, G., Lazzari, S., Ghys, P.D., Schwartlander, B. and Stanek, K.A. (2001) Epidemiological Analysis of the Quality of HIV Sero-Surveillance in the World: How Well Do We Track the Epidemic? *AIDS*, **15**, 1545-1554. <http://www.ncbi.nlm.nih.gov/pubmed/11504987>  
<http://dx.doi.org/10.1097/00002030-200108170-00012>
- [23] Toprani, A., Madsen, A., Das, T., Gambatese, M., Greene, C. and Begier, E. (2014) Evaluating New York City's Abortion Reporting System: Insights for Public Health Data Collection Systems. *Journal of Public Health Management & Practice*, **20**, 392-400. <http://www.ncbi.nlm.nih.gov/pubmed/24281129>  
<http://dx.doi.org/10.1097/PHH.0b013e31829c88b8>
- [24] Van Naarden Braun, K., Pettygrove, S., Daniels, J., Miller, L., Nicholas, J., Baio, J., Schieve, L., Kirby, R.S., Washington, A., Brocksen, S., *et al.* (2007) Evaluation of a Methodology for a Collaborative Multiple Source Surveillance Network for Autism Spectrum Disorders—Autism and Developmental Disabilities Monitoring Network, 14 Sites,

- United States, 2002. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, **56**, 29-40. <http://www.ncbi.nlm.nih.gov/pubmed/17287716>
- [25] Welk, G.J., Bai, Y., Saint-Maurice, P.F., Candelaria, N., Allums-Featherston, K.A. and Anderson, K. (2014) Design and Evaluation of the NFL PLAY 60 FITNESSGRAM Partnership Project. *Research Quarterly for Exercise and Sport*, in Press.
- [26] Zhu, W., Plowman, S.A. and Park, Y. (2010) A Primer-Test Centered Equating Method for Setting Cut-Off Scores. *Research Quarterly for Exercise and Sport*, **81**, 400-409. <http://dx.doi.org/10.1080/02701367.2010.10599700>
- [27] Boiarskaia, E.A., Boscolo, M.S., Zhu, W. and Mahar, M.T. (2011) Cross-Validation of an Equating Method Linking Aerobic FITNESSGRAM(R) Field Tests. *American Journal of Preventive Medicine*, **41**, S124-S130. <http://www.ncbi.nlm.nih.gov/pubmed/21961612>  
<http://dx.doi.org/10.1016/j.amepre.2011.07.009>
- [28] Saint-Maurice, P.F., Welk, G.J., Laurson, K.R. and Brown, D.D. (2013) Measurement Agreement between Estimates of Aerobic Fitness in Youth: The Impact of Body Mass Index. *Research Quarterly for Exercise and Sport*, **85**, 59-67.
- [29] Tomkinson, G.R. and Olds, T.S. (2007) Secular Changes in Pediatric Aerobic Fitness Test Performance: The Global Picture. *Medicine and Sport Science*, **50**, 46-66. <http://www.ncbi.nlm.nih.gov/pubmed/17387251>  
<http://dx.doi.org/10.1159/000101075>
- [30] Leger, L.A., Mercier, D., Gadoury, C. and Lambert, J. (1988) The Multistage 20 Metre Shuttle Run Test for Aerobic Fitness. *Journal of Sports Sciences*, **6**, 93-101. <http://www.ncbi.nlm.nih.gov/pubmed/3184250>  
<http://dx.doi.org/10.1080/02640418808729800>
- [31] (2012) *Fitness Measures and Health Outcomes in Youth*. The National Academies Press, Washington DC.
- [32] Corbin, C.B., Lambdin, D.D., Mahar, M.T., Roberts, G. and Pangrazi, R.P. (2014) Why Test? Effective Use of Fitness and Activity Assessments. In: *Why Test? Effective Use of Fitness and Activity Assessments*, The Cooper Institute.
- [33] Garcia-Calleja, J.M., Zaniewski, E., Ghys, P.D., Stanecki, K. and Walker, N. (2004) A Global Analysis of Trends in the Quality of HIV Sero-Surveillance. *Sexually Transmitted Infections*, **80**, i25-i30. <http://www.ncbi.nlm.nih.gov/pubmed/15249696>  
<http://dx.doi.org/10.1136/sti.2004.010298>
- [34] Armstrong, N. and Welsman, J.R. (2007) Aerobic Fitness: What Are We Measuring? *Medicine and Sport Science*, **50**, 5-25. <http://www.ncbi.nlm.nih.gov/pubmed/17387249>  
<http://dx.doi.org/10.1159/000101073>
- [35] Clark, R.G., Templeton, R. and McNicholas, A. (2013) Developing the Design of a Continuous National Health Survey for New Zealand. *Population Health Metrics*, **11**, 25. <http://www.ncbi.nlm.nih.gov/pubmed/24364838>  
<http://dx.doi.org/10.1186/1478-7954-11-25>
- [36] Mokdad, A.H., Stroup, D.F. and Giles, W.H. (2003) Public Health Surveillance for Behavioral Risk Factors in a Changing Environment. Recommendations from the Behavioral Risk Factor Surveillance Team. *Morbidity and Mortality Weekly Report. Recommendations and Reports*, **52**, 1-12. <http://www.ncbi.nlm.nih.gov/pubmed/12817947>
- [37] Dwyer-Lindgren, L., Freedman, G., Engell, R.E., Fleming, T.D., Lim, S.S., Murray, C.J. and Mokdad, A.H. (2013) Prevalence of Physical Activity and Obesity in US Counties, 2001-2011: A Road Map for Action. *Population Health Metrics*, **11**, 7. <http://www.ncbi.nlm.nih.gov/pubmed/23842197>  
<http://dx.doi.org/10.1186/1478-7954-11-7>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

