CITADEL CORRELATION ONE DATATHON

The London 2012 Olympics, Sports Participation, and Obesity: A Longitudinal Analysis Across Boroughs and Income Levels

George Hu[1] | **Kevin Huang**[2] | **Sid Rastogi**[3] | **Imad Rizvi**[4]

[1]Brown University, Providence, RI, 02912, USA

[2]Harvard University, Cambridge, MA, 02138, USA

[3]Harvey Mudd College, Claremont, CA, 91711, USA

[4]Yale University, New Haven, CT, 06520, USA

Correspondence
Email:
george_hu@brown.edu,
kevinhuang@college.harvard.edu
srastogi@hmc.edu
imad.rizvi@yale.edu

The UK Government and London originally included in its bid for the 2012 Summer Olympics the desire for the games to better the health of its residents, yet many studies have shown the health impacts of the Olympics to be negligible overall.

In our investigation, we focus on effects in specific regions and at specific income levels, using childhood obesity and adult sports participation as proxies for public health. We show that at the 95% significant level, a predominant majority of London's richest boroughs had increased sports participation and lower obesity levels after 2011. Furthermore, we show that the difference in playing sports between richer and poorer areas was exacerbated, along with moderate evidence that sports participation is a negative lagging indicator for childhood obesity certain wealthier regions.

1 | BACKGROUND

In the summer of 2012, Olympic fervor reached its height in London. The UK government had promised that hosting the Olympics would inspire widespread participation in sports, invigorating the populace and improving health conditions in the country for years to come. “We as a government will focus on [...] harnessing the United Kingdom’s passion for sport to increase grass roots participation [...] and to encourage the whole population to be more physically active,” wrote the Department for Culture, Media, and Sport as their first goal for the legacy of the Olympic Games, even before economic gain. However, studies and investigations since the Games have found that this goal has been far from achieved. A study by the International Academy of Sports Science and Technology [1] found little evidence of changes in physical activity in the short and long term, and other studies have found little effects from other major sports competitions as well [2][3].

Another trend found from studies on the Olympics are how the benefits may be conferred unequally upon the residents of the host country. Thousands of poorer residents may be misplaced, most employees hired for construction tend to already have jobs, and effects of tourism tend to have disproportionate impacts on wealthier areas of the city [4][5].

We sought to explore further the effects of the Olympics on health using data on obesity and sports participation in London before and after the Games. While it was unlikely that we would find an overall public health benefit as a result of the Games, we hoped to understand whether there were effects in certain regions and why certain London boroughs may have had more beneficial or adverse impacts. We were interested whether the unequal impacts of hosting the Olympics would manifest in public health improvements as well, and to understand how generalizable the impacts might be to other cities that have hosted or will host the Summer Olympics.

2 | DATA EXPLORATION

We set out a few main goals for our exploratory data analysis:

- Confirm that health and lifestyle indicators stayed relatively constant spanning the period of about five years before the London Olympics to 5 years after. This would be manifest in the proportion of London residents partaking in sports and the proportion that were overweight or obese.
- Find the extent to which there do exist regional differences in income across London boroughs.
- Obtain a basic understanding of the relationships between a region’s income level, sports activity, and obesity indicators.

2.1 | Datasets

For income data by borough, we used the provided *london_taxpayer_income* dataset. This provided central measures of individual annual income as provided by tax accounts. Since income is often right skew, we used the median individual income from this dataset.

For data on participating in sports, we used the provided *london_sports_participation* dataset, which supplied the percentage of residents in each borough of London that partook in no sports, one or more sports, and three or more sports per week. Each data point consisted of the above statistics for a given region of London or the UK in a given year, along with the area code and the population of the region.

For data on childhood overweight and obesity rates, we used an outside dataset titled "Prevalence of Childhood Obesity, Borough, Ward and MSOA", which is available from the London Department of Health [6]. The dataset includes the percentage of children in year 6 and reception (pre-school) who are overweight as well as the percentage who are obese, by bourough. In addition, it is longitudional, with annual data ranging from 2006 to 2018.

2.2 | Sports Participation Analysis

Reading the provided description of the sports participation dataset gives that the one or more category includes data from the three or more category. Thus, we had to subtract the 3+ percentage from the 1+ percentage in order to get the proportion in one to two sports so that our categories would be discrete and interpretable. In addition, the data included data from other regions across the UK outside of London, which was not our focus, so that data was filtered out.

We quickly found that for a given region and year, summing the three categories of 0, 1-2, and 3+ failed to result in 100 percent, indicating a level of non-response in the survey. Hence we normalized the percentages for each data point so that they summed to 100 to fix this, giving our desired visualization (Figure 1).

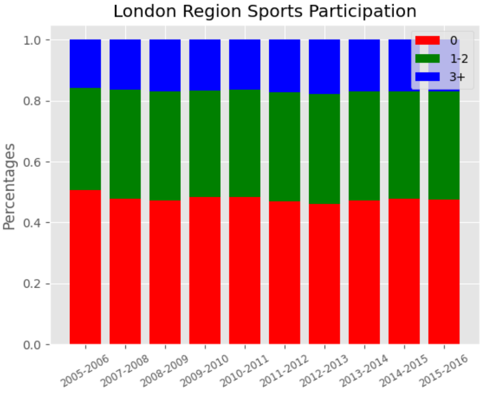


FIGURE 1 This graph represents the sports participation data of all of London over the time period 2005-2016. There do not seem to be any significant overall changes over time.

To aid with interpretation of sports participation, we also created a simple index that we will call the

Sports Participation Index (SPI), that is calculated as:

$$SPI = 0.5 * (\text{proportion playing 1-2 sports}) + (\text{proportion playing 3+ sports})$$

This way, an index of 0 would mean no residents are playing any sport, an index of 1 would mean all residents are playing three or more sports, and playing three or more sports is weighted twice just playing 1-2 sports. And using SPI, we can confirm that the population of London engaged in sports did not have any significant changes over the time interval of interest. Figure 2 displays the SPI of London as a whole over time.

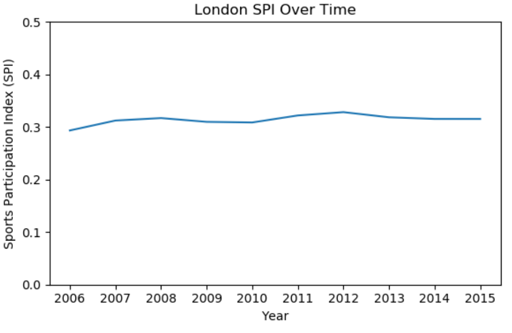


FIGURE 2 This graph represents the calculated values of the Sports Participation Index for London over time. The SPI for London remains relatively constant. There is a slight increase around the year 2012 that can likely be attributed to the Olympics, but this is still very insignificant.

2.3 | Overweight and Obesity Analysis

Next, we wanted to ensure claims that the benefit from the Olympics in terms of public health aggregated across London was negligible. We focused on obesity, specifically childhood obesity, as public health literature has shown that in developed nations, childhood obesity was a strong proxy for the general welfare of a region. Hence, our primary data point for public health outcomes in London was overweight and obesity statistics collected from children at the age of reception (equivalent to pre-school in the US, age 4) and at the age of Year 6 in primary school (age 12).

Similar to the sports participation data, we want to confirm that our childhood obesity measures have not trended significantly over the time period before and after the 2012 Olympics, as previously noted in literature showing negligible overall health impacts of the games. In the context of our data source, "overweight" is defined as overweight (BMI > 25), but not obese (BMI > 30), to distinguish the two groups, so we will use this definition as well throughout. Doing so, in figure 3 we look at the data for London as a whole.

Figure 3 shows that there are not many significant changes in proportion of children with unhealthy weight, aside from very mild trends of age 12 obesity increasing and the rest of the statistics decreasing by an extremely small factor. But these conflicting trends are indeed ambiguous in a holistic view of public health, so for visualization and analysis purposes, we created an index of overweightness and

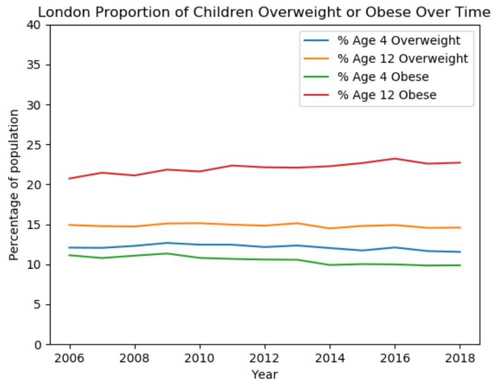


FIGURE 3 This graph illustrates the proportion of children categorized as overweight or obese in London. There do not seem to be significant changes in these percentages over time. Some mild trends can possibly be inferred, but any holistic trends are ambiguous.

obesity as well. The Unhealthy Weight Index (UWI) is calculated as:

$$\begin{aligned} UWI &= 0.002 * (\text{percentage age 4 overweight}) + 0.004 * (\text{percentage age 4 obese}) \\ &\quad + 0.003 * (\text{percentage age 12 overweight}) + 0.006 * (\text{percentage age 12 obese}) \end{aligned}$$

With this, an index value of 0 would mean that there are no overweight or obese children at age 4 and 12, and an index value of 1 would mean that all children at age 4 and 12 are obese. Obesity is weighed twice as much as overweightness, and age 12 is weighed 50% more than age 4, since literature has shown that obesity in adolescence carries over to adulthood much more frequently due to the presence of brown adipose tissue (more colloquially "baby fat") at age 4. Figure 4 confirms that the London UWI has negligible changes over time.

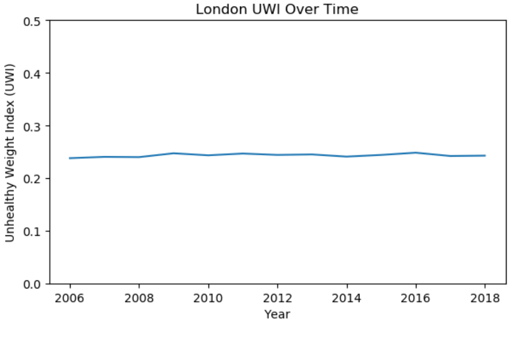
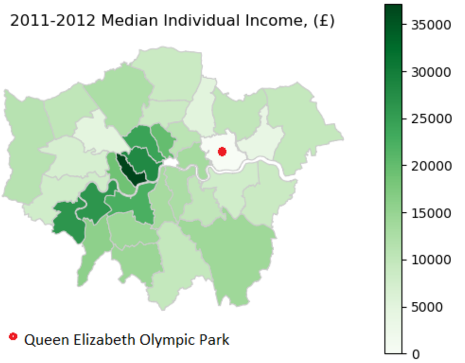


FIGURE 4 This graph illustrates the calculated value of the Unhealthy Weight Index (UWI). The UWI for London as a whole remained relatively unchanged over time.

2.4 | Income Analysis

Based on the data of taxpayer income by borough, focused on 2011-2012 before the tourism impacts of the Olympics should be realized, we plotted a graph of income by borough (ignoring City of London, which is not considered one of the 32 boroughs and is much smaller in population).



Borough	Income (£)
Kensington and Chelsea	37200
Westminster	34200
Richmond upon Thames	31500
⋮	⋮
Brent	20200
Barking and Dagenham	19600
Newham	17600

TABLE 1 Median individual incomes for London boroughs.

The wealthier boroughs seem to be more concentrated towards the center of the city and to its west/southwest. The wealthiest three boroughs measured by median income are Kensington and Chelsea, Westminster, and Richmond-upon-Thames, and the least wealthy boroughs are Newham, Barking and Dagenham, and Brent. Also, note that the Queen Elizabeth Olympic Park, which was the site of most of the 2012 Olympic events was built on the east side of London (in Newham), which is the relatively poorer side of London. In the next section, we find that this somewhat confounds our analysis that the London Olympics affected public health metrics based on regional wealth, because we found that how the London Olympics affected public health outcomes across regions depended on both wealth and proximity factors.

3 | STATISTICAL TESTING

In this section we use a variety of approaches to show that the Olympics caused noticeable changes in sports activity in some boroughs, which in turn caused noticeable changes in overweight and obesity. Furthermore, we aim to determine if regional income is a distinguishing factor for "some boroughs" as described earlier.

3.1 | The Relation Between the Olympics, Income, and Sports

Previous evidence indicates that there is an association between income and levels of exercise [7]. To confirm whether this trend holds in London, we ran three separate regressions of how median income predicts percentage of people exercising zero times per week, one-two times per week, and three or more times per week.

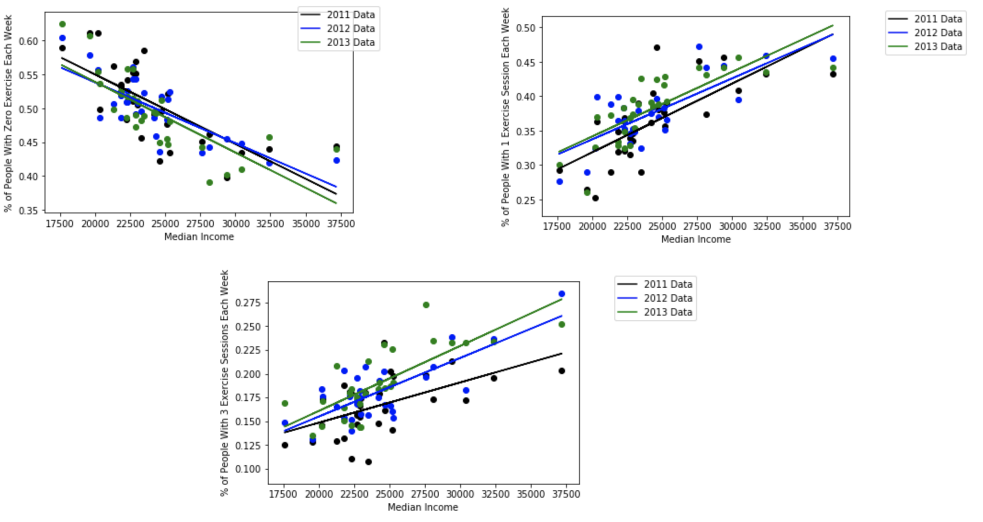


FIGURE 5 Various graphs with regression lines measuring relationship between income and sports participation at three time points. Top left is no exercise, top right is 1-2 sessions, and bottom is 3+ sessions per week. The inequality demonstrated on the bottom graph seems to increase over time.

For each of these comparisons, we used the 2011, 2012, and 2013 data. We were interested in exploring whether the slope of the relationship between median income and sports participation changed over the course of the Olympics. We found no significant changes in the zero exercise sessions per week and one-two exercise sessions per week group. However, in the category of people exercising three or more times per week, we found that the slope of the regression line and the strength of the relationship between median income and sports participation tended to increase between 2011-2013, which is visualized in Figure 5 and described in Table 2.

Year	Zero Exercise Group	1-2 Exercise Group	3+ Exercise Group
2011	-1026 (0.513)	997 (.522)	423 (.301)
2012	-897 (.608)	886 (.584)	616 (.588)
2013	-1042 (.585)	937 (.610)	684 (.522)

TABLE 2 Each element of the table is the slope of the regression in scaled units, along with the strength of the regression in parentheses.

Given that we found evidence of unequal impacts on various boroughs based on income, we conducted hypothesis testing on specific boroughs. Intending to look for lasting effects from during and after the Olympics, our data of interest was data from 2011-2012 until 2015-2016. Thus we have the following hypotheses for a given London borough:

- H_0: The sports participation rate in a borough after 2012 was the same or less than the amount before.
- H_a: The sports participation rate in a borough after 2012 was more than the amount before.

We measure sports participation using SPI, and since we are unsure of the natural distribution of SPI, and have a somewhat limited sample size, we proceed using a permutation test on the sample means of SPI before and after 2012 for each borough. The results of the test can be visualized in Figure 6. At

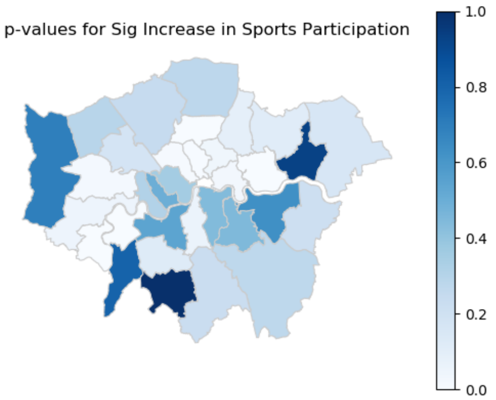


FIGURE 6 p-values for the permutation test showing low p-values and thus increases in sports for much of northern and western London. This northern portion can mostly be explained by proximity to Queen Elizabeth Olympic Park.

the 95% significance level, the boroughs with significant p-values are (p-value, median income rank in parentheses): Camden (p = 0.024, 4th), Ealing (p = 0.032, 28th), Hackney (p = 0.047, 17th), Haringey (p = 0.012, 25th), Islington (p = 0.016, 6th), Newham (p = 0.012, 32nd), Richmond upon Thames (p = 0.012, 3rd), and Tower Hamlets (p = 0.036, 12th).

Hackney, Haringey, and Newham are all in proximity of Queen Elizabeth Olympic Park, which staged the main events of the 2012 Summer Olympics, so their anomalous increases in sports can be attributed to that as a main factor. Hence, out of the remaining five boroughs that showed significant increases in sports participation, a staggering four of them are in the top half of boroughs for median income, and even three of them in the top quartile. Therefore, income level and general demographics are likely to play a large role in having the Olympics influence sports participation.

3.2 | Adding in Overweight and Obesity

Another important facet of public health is the overweight and obesity rates. We also tested whether there was a significant effect from the Olympics on lowering the overweight and obesity rates, as measured through the Unhealthy Weight Index (UWI). Our two periods of interest were the period before 2012 (from 2006 to 2012) and the period after 2012 (from 2012 to 2018), and as before, we are interested in unequal impacts on various bouroughs. Therefore we have the following hypotheses for a given London borough:

- H_0: The UWI in a borough after 2012 was the same or greater than the amount before.
- H_a: The UWI in a borough after 2012 was less than the amount before.

To test this hypothesis, for each borough, we conducted an one-sided t-test for the two groups of data (before and after 2012), assuming that the variance for each of the two groups was equal. Then, the results of the test, as well as the regions in which we have a 95% significance level (p-value < .05), are visualized in Figure 7. At the 95% significance levels, the bouroughs with significant p-values are (p-value,

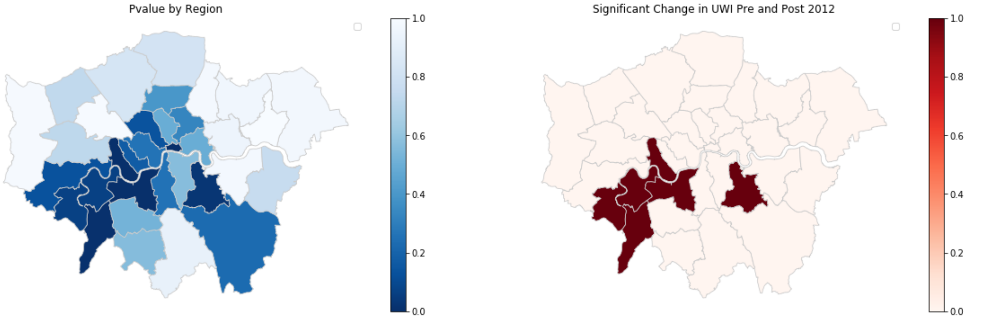


FIGURE 7 p-values for the permutation test showing low p-values and thus decreases in UWI for much of western London. We find significance for five bouroughs, mainly located in the west of London.

median income rank in parenthesis): Kingston Upon Thames (p = .002 ,8th), Wandsworth (p = .003, 5th), Richmond Upon Thames (p = .032, 3rd), Hammersmith and Fullham (p = .002, 7th), Lewisham (p = .012, 21st). Aside from Lewisham, all of the boroughs which experienced a significant decline in the UWI following the 2012 Olympics were in the top quarter of London boroughs by income, which further supports our hypothesis that a bourough's income level is likely to play a significant role in the effect of Olympics on public health outcomes. Since the majority of the bouroughs here are located on the West side of London, while the Queen Elizabeth Olympic Park where the main events of the 2012 Summer Olympics took place are located in the East of London, we can rule out proximity effects to explain the Olympics's disproportionate effect on overweight and obesity rates between regions.

3.2.1 | Linking Sports Participation to Public Health

If we take a look at some of the time series for SPI and UWI in boroughs that we are primarily interested in (increase in SPI, decrease in UWI, and wealthy), an interesting trend emerges. For example, take Islington, as in Figure 8.

It appears that there is a time dependent negative correlation between SPI and UWI. To test this,

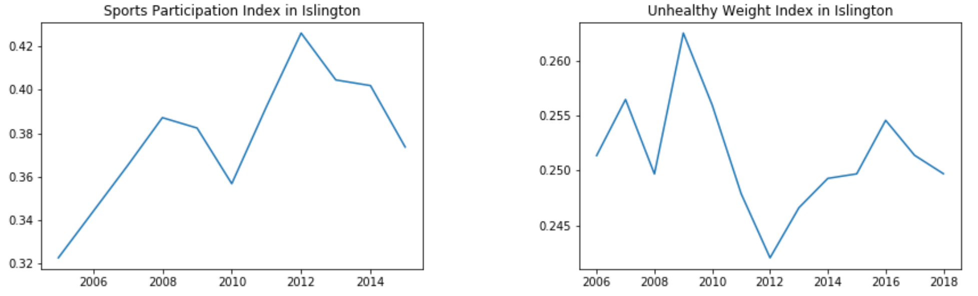


FIGURE 8 *Left:* SPI for Islington, *Right:* UWI for Islington. The graphs seem somewhat inverted with each other.

we look at the cross correlation between the two time series over the common time period 2006-2015 represented in both graphs. By analyzing the cross correlation, we can determine if there are any time lagged correlations between the two, as in Figure 9.

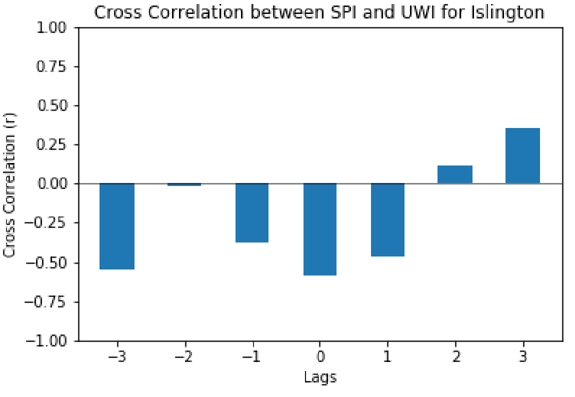


FIGURE 9 Cross Correlation plot for SPI and UWI. Lag time steps are in years, and lag k means you are considering the correlation between t-indexed time series SPI_t and UWI_{t-k}

From Figure 9, we get that the cross correlation crosses a mild significance threshold of $|r| > 0.5$ at lag 0 and lag -3. This means that the SPI is negatively correlated with the UWI for the same year, as well as the UWI three years later. Hence, this evidence is strongly suggestive that SPI can act as a negative preceding indicator for UWI, which matches our expectation; participating in more sports leads to a decreased prevalence of obesity. This claim is further supported by the cross correlation plots of Camden and Richmond Upon Thames. Note that such boroughs are not cherry picked; they are specifically selected since they displayed a significant change in SPI attributed to the Olympic Games, without having the trait of proximity to the main venues.

For the case of Camden, only lag negative one passes the $|r| > 0.5$ threshold of significance, indicating that UWI is a inverse lagging indicator by one year for SPI. And for more mild correlations $0.4 < |r| < 0.5$,

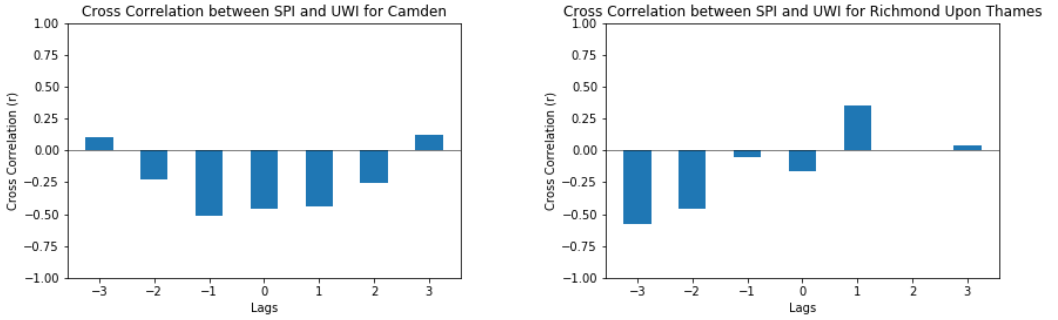


FIGURE 10 *Left:* Cross Correlation for Camden, *Right:* Cross Correlation for Richmond Upon Thames. The plots both show mild to moderate negative correlations.

lags 0 and 1 both qualify, indicating a general negative correlation between the two.

The time stepped correlation is more evident in the data for Richmond Upon Thames. Here only lag -3 is significant, giving that a change in SPI at year t is likely to correspond with an opposite change in UWI at year $t+3$. Lag -2 has a mild correlation as well, providing credence to the idea that SPI and UWI are negatively correlated in wealthier boroughs, with a possible lagged delay in effects for UWI.

4 | CONCLUSION

We provided evidence supporting our hypothesis that impacts of the Olympics are distributed unequally among residents of the host city. Like many effects of the Olympics, it seems that benefits in public health occur predominantly for wealthier residents of the host city, along with those situated nearest to major stadiums. Another observation is that most of these differences were predominantly in the category of residents participating in 3 or more sports a week; those in wealthier areas were relatively more likely than their counterparts to participate in extensive exercise in the year after the Olympics compared to the year before. Though we could not find relevant health data split by region from other host cities, future data for the Tokyo Olympics should be explored to determine whether this trend holds in general. Another area to explore would be to examine data from further after the Olympics to determine how much of a long-term impact they have on obesity and exercise.

Our findings also indicated sports participation as a potential lagging indicator for obesity, as we assume childhood obesity and adult sports participation to be relevant proxies for overall obesity and exercise in general. In general, it seems that while the Olympics have the potential to promote exercise and healthier lifestyles in host cities, it will require extra consideration from the government to ensure these effects are gained equally — more focus on less wealthy regions along with areas further from the center can potentially help improve overall health in an Olympic host city.

References

[1] Engalycheva K, Chappelet JL. Sport Participation in Host Countries before 89 and after the Olympic Games: Do the Games Change Something? Collected Insights from the Field of Sport, Volume 2: Sustainability and Legacy 2015;p. 89–108.

[2] Veal AJ, Toohey K, Frawley S. The sport participation legacy of the Sydney 2000 Olympic Games and other international sporting events hosted in Australia. Journal of policy research in tourism, leisure and events 2012;4(2):155–184.

[3] Coalter F. London 2012: a sustainable sporting legacy. After the Goldrush: a sustainable Olympics for London London: ippr and Demos 2004;.

[4] Worles K. The Dark Side of Hosting the Olymipcs. Inequalityorg 2017;.

[5] Zimbalist A. Circus maximus: The economic gamble behind hosting the Olympics and the World Cup. Brookings Institution Press; 2020.

[6] of Health LD. Prevalence of Childhood Obesity, Borough, Ward and MSOA. London Opinion Research and General Statistics;.

[7] Kosteas VD. The effect of exercise on earnings: Evidence from the NLSY. Journal of Labor Research 2012;33(2):225–250.