

COVID-19 Diffusion Before Awareness: The Role of Football Match Attendance in Italy

Journal of Sports Economics
2022, Vol. 23(5) 503-523
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15270025211067786
journals.sagepub.com/home/jse



Vincenzo Alfano^{1,2} 

Abstract

Anecdotal evidence suggests that football matches may have played a role in the spread of COVID-19 all over Europe. Nevertheless, from a scientific point of view, the impact of football matches on the spread of COVID-19 remains unclear. In this paper we study, via a quantitative analysis, the case of Italy, a country badly affected by COVID, and one where attending football matches is very popular. We consider the impact of matches played in January and February 2020 on the dynamic of the pandemic in March and April the same year. Our results, which consider all levels of Italian professional football, and the highest level of amateur football, show that matches played in January and February had an impact on the evolution of the pandemic in March and April. These results suggest that great care must be taken before considering re-opening stadia.

Keywords

football matches, outdoor gatherings, mass events, COVID-19

JEL Codes: I18, H12, Z28

¹ Department of Economics, Westminster International University in Tashkent, Tashkent, Uzbekistan

² Center for Economic Studies of Munich, CES-ifo, Munich, Germany

Corresponding Author:

Vincenzo Alfano, Department of Economics, Westminster International University in Tashkent, Tashkent, Uzbekistan; Center for Economic Studies of Munich CES-ifo, Munich, Germany.

Email: valfano@wiut.uz

Introduction

The Italians lose wars as if they were
 football matches and football
 matches as if they were wars.
 - Winston Churchill

Football¹ plays a very important role in Italian culture: the famous quote from Winston Churchill with which we opened this section remains valid today. It is by far the favorite sport of Italians, and supporting the local team is a very common hobby, one that is sometimes experienced as a duty. Italians continue to attend football matches in person, and fans often gather to support their favorite team. After a number of years in which live attendance of football matches was in decline, in more recent years Italy has seen a resurgence of the number of fans going to stadiums (Figure 1). Average attendance of Serie A games surpassed 25,000 fans in 2018/2019 for the first time since the 2009/2010 season, and has continued to grow in subsequent years, reaching 27,226 in the 2019/2020 season.

The recent pandemic has affected this habit significantly. With the discovery of the first cases of a new strain of coronavirus (COVID-19) at the end of January

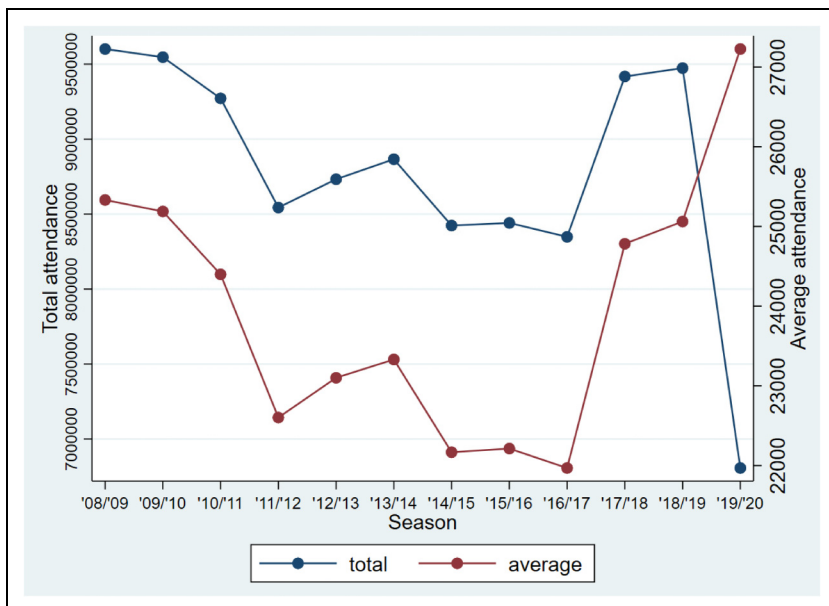


Figure 1. Total (left handed scale) and average (right handed scale) attendance from Serie A 2008/2009 to 2019/2020 seasons.

2020, and the worsening of the situation that rapidly ensued, Italy (the first country in the Western world to be severely affected by the pandemic) saw the start of a cycle of policies and attempts by the government to stop the spread of the virus. On 1 March 2020, with his second decree regarding the pandemic emergency, Dr. Giuseppe Conte, the President of the Italian Council of Ministers, banned attendance of all sports events, among other measures. As it became increasingly clear that the situation in Italy was rapidly deteriorating, President Conte issued another decree just three days later, on 4 March, closing all stadiums, thus forcing *Lega Calcio* to suspend the Italian football league (together with other sports).

Although the first cases reported in Italy dated to 31 January 2020, when a Chinese couple (originally from Wuhan) traveling in the country were tested positive in the Spallanzani National Institute for Infectious Diseases, with the first autochthonous contagion being reported only on 24 February, evidence later emerged that the virus had been circulating in the Italian peninsula since at least December 2019. One study (La Rosa et al., 2020) found traces of the virus in wastewater samples in Turin and Milan, taken in the purifiers of urban centers, and used these as evidence of the circulation of the virus in the population.

Thus, given that the virus was circulating long before the authorities began to trace its movements, we may consider it likely that mass events contributed to the spread of the disease. At the same time, Reade and Singleton (2020) report that the pandemic has had negative effects on the demand for football attendance in Italy, and Reade et al. (2020) discover that in the only professional league that was not suspended during the COVID-19 pandemic, the fans spontaneously self-distanced (Reade et al., 2020). Thus, from a theoretical point of view, the effect of football matches on the COVID-19 dynamic remains unclear.

It has been suggested that football played a major role in spreading the contagion. In particular, several articles in the press have suggested that a Champions League match between Valencia and Atalanta (a team from the city of Bergamo, one of the epicenters of the pandemic in Italy and the most severely affected province in the first COVID-19 wave) was one of the reasons behind “Italy’s coronavirus disaster”, as the *Wall Street Journal* called it.²

More generally, was the news of the existence of a virus and self-control of the fans enough to prevent football matches from acting as spreader events? Or did football matches contribute to the spread? On the one hand, football matches are hosted outdoors, but on the other they often see many people gathered in a restricted space, and, as highlighted by Olczak et al. (2020), these tend to group together even if the stadium is not full, and chant to encourage their team, thus favoring the spread of saliva droplets, the main vehicle through which the virus travels (Han & Ivanovski, 2020). In other words, did the concentration of people in leisure events such as football matches have major public health consequences?

Surprisingly, as noted by Sassano et al. (2020), the literature has not yet convincingly answered this question. After several studies devoted to measuring the impact of non-pharmaceutical interventions on COVID-19 diffusion (Alfano & Ercolano,

2020; Cruz & Dias, 2020; Figueiredo et al., 2020; Lau et al., 2020), the present article, following the pioneering approaches of Olczak et al. (2020) on England and Fischer (2021) on Germany, aims to shed some light on the issue with regard to Italy, via a quantitative approach. It is important to specify that Italy has been a very special case so far as COVID-19 is concerned. After an initial phase in Europe when a few small clusters of the virus were localized in Germany, France and the UK, a major outbreak erupted in northern Italy. The entire area, and in particular the regions of Lombardy, Emilia Romagna, and Veneto, was severely affected by COVID-19. Starting on 23 February, 10 municipalities in Lodi province (in the region of Lombardy) and one in Padua province (in the region of Veneto) were locked down (Riccardo et al., 2020). On 11 March 2020, the central government decided to extend the lockdown across the entire country, a measure that was put into action the very next day. For all these reasons Italy may be considered the first Western democracy to have had to seriously confront the COVID-19 pandemic.

We consider the impact of sport events on the coronavirus pandemic to be a very important issue, given policymakers' unavoidable decision about whether to close stadiums early during a pandemic to reduce the spread of the virus, and the debate as to whether this justifies the economic costs that early closure entails. The optimal tradeoff between benefits and risks may be found more easily using the results of an empirical analysis aimed at assessing the impact of sports events on the spread of COVID-19. This is the aim of the present work.

To the best of our knowledge, no article has yet focused its empirics on the Italian case with regard to the spread of airborne viruses in mass outdoor events. This is surprising, since Italy seems to be the best possible setting in this regard, for two main reasons. The first is that, by the time public attendance of football matches was suspended, the virus was much more prevalent in Italy than in any other Western country. Second, unlike most other Western countries, Italy has only one dominant sport, namely, football, in terms of both popularity and match attendance. According to Boeri and Severgnini (2014), three Italians out of four are interested in the Serie A, and about 600,000 regular football games are played in Italy every year, more than in the UK, for instance. Over 10% of the Italian male population plays soccer regularly (ISTAT, 2005), and the national team has about 28 million supporters. This makes it unlikely that other sport events have had a comparable impact on this dynamic.

The importance of this topic is confirmed by the attention that the literature has devoted to the question elsewhere. Previous research suggests that mass events act as 'super-spreaders' of airborne viruses. This is the conclusion reached by Stoecker et al. (2016) with regard to the impact of American football on influenza-caused deaths, and by Cardazzi et al. (2020) with regard to professional sporting events on influenza mortality. Similar early results have been reached with regard to COVID-19 in Ahammer et al. (2020), which studies the case of indoor events in the USA, and in a study mentioned previously whose interests are closer to our own, the paper by Olczak et al. (2020), which, studying the impact of football

events in England, discovers that matches had an impact on COVID-19 cases, deaths and excess deaths. Also relevant is the work of Fischer (2021), which studies the impact of German football events on the second COVID-19 wave in Germany, finding that they did contribute to an increase in cases. It is also important to highlight that we should not expect any impact in this mechanism due to different government efficiency (as suggested by Alfano & Ercolano, 2021a), nor to differences in social capital stocks in the population (Alfano & Ercolano, 2021b), given that at this time non-pharmaceutical interventions were not in place in Italy.

Following previous literature (Olczak et al., 2020), we built a dataset of football match attendance in the various Italian provinces in January and February, in order to study the impact of this on the evolution of the pandemic in the subsequent months. While at first glance it may seem unlikely that an impact will be found within such an extended timeframe, given that 97.5% of people who develop symptoms do so within 11.5 days of infection, which implies a confidence interval of between 8.2 and 15.6 days (Lauer et al., 2020), it is important to note that in March and April there was not as much awareness as there is now about asymptomatic and paucisymptomatic patients. This, as already highlighted by Brown (2020), biased the statistics with which the government monitored the situation, and contributed greatly to the spread of the virus. It also seems to imply that among the people who were infected in January and February, possibly because of their attendance of mass gathering events such as football matches, many may have been asymptomatic and paucisymptomatic, and were thus unaware of their condition, which would have contributed to them spreading the virus in subsequent months (Nicastri et al., 2020). The effects of these events would therefore become evident only months later, when a proper testing policy had been adopted, and more appropriate data about the spread of the coronavirus had emerged.

Using both Ordinary Least Squares (OLS) and Fractional Probit (FP) estimators, we found empirical evidence that the provinces that hosted more football matches in January and February 2020 were more severely affected by COVID-19 in March and April of the same year. This finding is robust to a number of controls and different estimators.

The remainder of the paper is organized as follows. After this introduction, the next section describes the methodology adopted and the data used; section three discusses the results, while the final section concludes.

Methodology and Data

In order to estimate the impact of football matches on the spread of COVID-19 in a subsequent month, following Olczak et al. (2020), we model the problem with this equation:

$$C_{pt} = \alpha + \beta_1 M_{pt-1} + \beta_2 X_p + \varepsilon \quad (1)$$

where C_{pt} is the number of per capita COVID-19 cases in province p at time t ; M_{pt-1} is a measure of the impact of football matches on C_{pt} (which can be either the per capita number of football events held in province p at time $t-1$, or the per capita attendance of all the football events held at time $t-1$ in province p); X_p is a matrix composed of a set of time-invariant controls, to avoid a biased estimate of the impact of M on C due to omitted variables; and finally ε is, as usual, the error term. Please note that, as highlighted by Olczak et al. (2020), when involving time dynamics, due to the lag of the independent variable M with respect to the dependent variable C , the equation can be estimated with cross-sectional data, and does not have a time dimension in the data.³

As is clear from equation (1), β_1 measures the impact of the football match proxy M on COVID-19 spread C . In order to estimate this equation empirically, we need to operationalize a proxy of coronavirus spread in Italian provinces, a proxy of football matches and their attendance, and finally a number of controls to take into account possible concurrent and alternative cases of the spread of infection.

With regard to the first proxy, we operationalized C_{pt} as the variable *Cases*, obtained from the Italian Ministry of Health dataset on COVID-19 cases for each Italian province. This is included from 24 February, when the series begins with the first local case reported in Italy, to 31 April, when we decided to close the series, since it was unlikely that we would see any effect due to stadiums, given that they had already been closed for two months. It reports COVID-19 cases day by day and province by province. We transformed these data into per capita terms by dividing them by the provincial populations in 2020 (data from ISTAT), in order to avoid biases due to the different sizes of the provinces, and to facilitate interpretation of the coefficients.

Unfortunately, other measures of the effect of COVID-19, such as deaths, are available for Italy at a regional but not provincial level. Given that Italy is divided into 20 regions and (for statistical purposes) 107 provinces, in order not to lose too many observations (and thus statistical power) in our cross-section, we chose to focus our analysis on the spread of coronavirus. This of course means that, unlike other works on the subject such as that of Cardazzi et al. (2020), Olczak et al. (2020), or Fischer (2021), our analysis will not determine the impact of football events on COVID-19-related deaths, but only on the spread of cases. While we recognize that this is a potential source of bias, since the quality of the estimation of cases has been questioned (ISTAT and Italian Ministry of Health, 2020), it is important to highlight, on the other hand, that the reported cases should be tightly correlated with real ones, and thus the potential bias may affect the magnitude of the estimate, but not its statistical significance. Nevertheless, we urge the reader to take our results with caution, given this potential flaw. Since the series began on 24 February, we summed up the daily cases in monthly totals per province for the months of March and April, obtaining two variables, *Mar Cases pc* and *Mar-Apr Cases pc*, alternative operationalizations of our dependent variable C_{pt} .

The second proxy, M , regarding the impact of football events on COVID-19, is operationalized in two different ways. M is either the per capita total number of football events that have happened in province p at time t or the per capita total number of supporters present at these events. As regards which events we considered, we gathered data (from <https://www.stadiapostcards.com/archivio.htm>, a website collecting data on attendance in Italian matches) about all the levels of Italian professional football (and thus about *Serie A*, *B*, and *C*) and also about the international matches of the Champions and Europa Leagues hosted in Italy, as well as matches in the Italian Cup (*Coppa Italia*). In short, these are all the professional football matches played in the first quarter of 2020, before the ban on attendance that was implemented on 1 March. Attendance is computed from this source with the sum of tickets sold and subscribers. While it may be that this operationalization underestimates the actual number of fans attending, given the possibility of people buying tickets without actually attending a match, it is important at the same time to highlight that press accreditation, free tickets and similar giveaways are not computed in this source, and are thus likely to balance the operationalized attendance with what is reported.

In order to expand the analysis, we also considered the next level in Italian football, the highest amateur-level competition: the so-called *Campionato Dilettanti – Serie D*. Unfortunately, with regard to this level, precise data on attendance for these games is not available in a disaggregated per match form, but only as an average (on <https://www.transfermarkt.it/>, a website collecting data about football teams). Considering that there is lower attendance at this level (the average was 606 spectators in January and 639 in February), and that (unlike in *Serie A*) there are no matches against very famous teams that may result in unusually high attendance (such as when a minor team in *Serie A* plays against Juventus or another top-tier team), we expect a lower variance from this variable, and for this reason we believe that the average is an acceptable proxy in order to estimate the attendance of such football events. All these data have been transformed into their per capita equivalents, to facilitate the readability of the coefficients, by dividing them by the populations of the provinces in 2020 (data from ISTAT).

Table 1 summarizes the number of matches played per month, and the league or cup they were played for. As can be seen, 545 matches were played in January, and 525 in February, giving a total of 1,070 matches in the two months. 98 of the 107 Italian provinces hosted at least one match in these two months: we proceeded to the listwise deletion from the dataset of provinces that did not host football events in January or in February.⁴ The total numbers of matches and attendees were summed in order to compute two proxies of M : *Matches*, the total number of matches per capita hosted per province p in each month t , and *Public*, the total number of attendees per capita per province and month.

Finally, regarding the controls included in the X matrix, the variables included are as follows. First of all, we control for the shares of population by different age bands (*ShPop 20-30*, *ShPop 30-40*, *ShPop 40-50*, *ShPop 50-60*, *ShPop 60-70*, *ShPop over 70*⁵) in each province, since it is important to control for different sizes in terms of the

Table I. Matches Included in the Dataset.

League or Cup	January	February	Total
Serie A	40	39	79
Serie B	21	48	69
Serie C – Group A	40	37	77
Serie C – Group B	40	37	77
Serie C – Group C	40	51	91
Champion League	0	2	2
Europa League	0	1	1
Coppa Italia	12	2	14
Serie D – Group A	36	35	71
Serie D – Group B	50	30	80
Serie D – Group C	50	33	83
Serie D – Group D	36	31	67
Serie D – Group E	36	36	72
Serie D – Group F	36	35	71
Serie D – Group G	36	36	72
Serie D – Group H	36	36	72
Serie D – Group I	36	36	72
Total	545	525	1,070

inhabitants of the provinces, given that age plays a role in the vulnerability to COVID-19, and hence in the degree of spread and the number of coronavirus cases (Davies et al., 2020).

It is important to note that the literature has highlighted how COVID-19, because of its epidemiological characteristics, is easily spread in mass gatherings (Riccardo et al., 2020), and thus population density seems to be a very important determinant of diffusion within a province. We control for this in two ways, one of which is more direct and one of which is indirect. First, we include among our controls in X the population density (*PopDens*) of the province, as a straightforward operationalization of the population density. Second, given that Italian provinces are highly heterogeneous (Ercolano, 2012), following Alfano et al. (2021) we include among the controls another variable (*ShU50k*) that signals the share of municipalities in the province that have a population below 50,000 inhabitants, to take into account not only the average provincial population density, but also these differences among the Italian provinces. This is also likely to play a role in the other opportunities of spreading the virus in the province, such as the availability of public transportation, and other characteristics linked with the structural features of the province that are likely to influence the spread of COVID-19 after an initial contagion. As highlighted by Alfano et al. (2021) with regard to opening schools, the structure of a province (on an ideal continuous line that goes from provinces constituted by many small municipalities to those characterized by a few big cities) may play an important role in

favoring the spread of COVID-19. Indeed, after an initial contagion event, the presence of large supermarkets or public transportation in bigger cities may easily transform an asymptomatic (or paucisymptomatic) football fan into a coronavirus super-spreader. All the data discussed so far are gathered from the Italian National Institute of Statistics (ISTAT).

Furthermore, given that income has been suggested as playing a role in the likelihood of becoming infected (Elgar et al., 2020), and given that it is a good proxy to control for other unobservable determinants, such as education, the culture of prevention, and the possibility of taking precautions, we have decided to include it (*Income*), operationalized as the mean taxable income per province (data from Italian Ministry of Economic and Finance).

Finally, given the exponential nature of COVID-19 contagion, the total number of cases of coronavirus in February, i.e., the month before the one computed in the dependent variable, is included, in order to control for other underlying variables affecting monthly COVID-19 cases. Descriptive statistics of the variables included in the model are presented in Table 2.

Results

In order to estimate equation (1) empirically, we first run a baseline regression through an OLS estimator, weighting the observations for the population of the province. Given the exponential nature of the contagion, we expect to have more significant spread in more populated provinces. In the absence of endogeneity, the best linear unbiased estimator of the coefficients is given by the ordinary least squares estimator, provided it exists (Neyman, 1934). Our regressions, performed with robust standard errors, could nonetheless be affected by this potential source of bias, since both Reade and Singleton (2020) and Reade et al. (2020) have found that football attendance is influenced by COVID-19 cases. Moreover, in the second half of February some football matches were postponed as a precaution, and thus the second operationalization of M could also be affected by this bias. Nonetheless, it is important to highlight that by including the total number of cases in the previous month among the regressors, we should reduce this bias by capturing the variance due to this mechanism in that independent variable, thus removing this bias from the independent variables of greatest interest.

The results are presented in Table 3. As already explained, we regressed the number of COVID-19 per capita cases for each Italian province in March on the number of football matches per capita hosted in January and February (3.1); and on the total per capita attendance of these matches (3.2); and similarly the number of COVID-19 per capita cases by province in March and April on the per capita number of football matches in January and February (3.3), and on the total per capita attendance of these matches (3.4); the regressions of course also include all the controls previously discussed.

Table 2. Descriptive Statistics.

Variable	Label	Obs	Mean	Std. Dev.	Min	Max	Source
Share of provincial population between 20 and 30 years old	ShPop 20-30	98	0.100	0.009	0.083	0.124	ISTAT
Share of provincial population between 30 and 40 years old	ShPop 30-40	98	0.113	0.007	0.091	0.130	ISTAT
Share of provincial population between 40 and 50 years old	ShPop 40-50	98	0.148	0.006	0.134	0.165	ISTAT
Share of provincial population between 50 and 60 years old	ShPop 50-60	98	0.158	0.005	0.144	0.170	ISTAT
Share of provincial population between 60 and 70 years old	ShPop 60-70	98	0.126	0.007	0.107	0.146	ISTAT
Share of provincial population over 70 years old	ShPop over 70	98	0.180	0.020	0.127	0.230	ISTAT
Population density in the province	PopDens	98	281.935	399.203	35.748	2,632.711	ISTAT
Mean taxable income in the province	Income	98	17,870.9	2,971.799	12,867.61	24,778.33	Italian Ministry of Economic and Finance (MEF)
Share of municipalities belonging to the province with a population under 50,000 inhabitants	ShU50k	98	0.719	0.205	0.128	1	Authors calculation from ISTAT data
Total number of matches played in the province in January and February per capita	Jan Feb Matches pc	98	10.918	8.475	3	45	Stadiopostcard and Transfermarkt

(continued)

Table 2. (continued)

Variable	Label	Obs	Mean	Std. Dev.	Min	Max	Source
Total number of people attending football matches in the province in January and February per capita	Jan Feb Public pc	98	43,564.47	102,944.7	1,000	761,951	<i>Stadiapostcard and Transfermarkt</i>
Total number of COVID-19 cases in the province in February per capita	Feb Cases pc	98	30.571	108.877	0	831	<i>Italian Ministry of Health</i>
Total number of COVID-19 cases in the province in March per capita	Mar Cases pc	98	11,466.99	20,417.29	87	123,346	<i>Italian Ministry of Health</i>
Total number of COVID-19 cases in the province in March and April per capita	Mar-Apr Cases pc	98	58,904.93	89,617.16	1,452	539,320	<i>Italian Ministry of Health</i>

Table 3. OLS: Determinants of Monthly COVID-19 Cases.

	(3.1)	(3.2)	(3.3)	(3.4)
	March Cases pc	March Cases pc	Mar-Apr Cases pc	Mar-Apr Cases pc
ShPop 20-30	-0.753 (-1.24)	-0.804 (-1.23)	-1.970 (-0.77)	-2.211 (-0.83)
ShPop 30-40	-0.890*** (-2.92)	-0.796* (-1.83)	-3.516*** (-2.82)	-3.363* (-1.94)
ShPop 40-50	-0.0741 (-0.19)	-0.0455 (-0.11)	-0.985 (-0.49)	-0.897 (-0.43)
ShPop 50-60	-1.207*** (-2.71)	-1.211** (-2.53)	-4.418** (-2.63)	-4.492** (-2.52)
ShPop 60-70	-0.486 (-1.12)	-0.449 (-1.05)	-1.463 (-0.87)	-1.275 (-0.77)
ShPop over 70	-0.374 (-1.56)	-0.318 (-0.97)	-1.164 (-1.16)	-1.066 (-0.84)
PopDens	-0.00000503** (-2.29)	-0.00000450* (-1.68)	-0.0000224** (-2.34)	-0.0000213* (-1.98)
Income	0.00000165 (1.55)	0.00000136 (1.23)	0.0000144*** (3.43)	0.0000132*** (3.12)
Share of pop. in mun under 30k inhab	0.0188 (1.48)	0.0237* (1.67)	0.0745 (1.51)	0.0951* (1.77)
Feb Cases	0.000209*** (5.65)	0.000212*** (5.43)	0.000590*** (4.35)	0.000596*** (4.18)
Jan-Feb matches pc	252.9** (2.37)		929.0** (1.99)	
Jan-Feb public pc		0.00360 (0.10)		0.0447 (0.31)
Constant	0.475** (1.99)	0.459* (1.67)	1.600* (1.67)	1.589 (1.49)
Observations	98	98	98	98
R ²	0.743	0.736	0.713	0.707
Adjusted R ²	0.711	0.702	0.677	0.669

t statistics in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Robust standard errors, Observations weighted by province population size.

Results suggest that the per capita number of hosted matches in January and February had a statistically significant impact on the per capita number of coronavirus cases both in March and in March and April. On the other hand, the per capita attendance of these matches does not have a statistically significant impact on the number of COVID-19 cases. The first operationalization of M is statistically significant at 5% in explaining the per capita COVID-19 cases both in March and in March and April. We consider this to be a very important threshold, especially given the reduced number (98) of observations.

So far as the control variables are concerned, as expected, the cases in the previous month (Feb cases) have much of the statistical significance, and so does *Income* for specifications 3.3 and 3.4, which refer to cumulative March and April cases. This is not unexpected, given, as we said, the great amount of information that this variable brings. It has, as expected, a positive sign, suggesting that provinces with more cases in February and the richer provinces have more COVID cases in, respectively, March and March and April. Richer provinces, as explained previously regarding the relevance of income to many unobserved variables, are also those populated by citizens that are more likely to travel, whether in Italy or abroad, and attend more public events.

At the same time, *ShU50k* also shows a certain degree of statistical significance in some specifications (3.2 and 3.4), and surprisingly has a positive sign. The provinces with a bigger share of their population in tiny municipalities are more affected by COVID-19. This result is further confirmed by the coefficient of *PopDens*, which has in all the specifications a negative and statistically significant sign. While this may seem surprising in light of the considerations made so far with regard to population density and its impact on the spread of COVID-19, we propose the following interpretation of these results. It is possible that provinces in which a large share of the population lives in tiny municipalities, and those with low population density, had (before greater public awareness of COVID-19 and the lockdowns of March) more of their citizens moving toward bigger cities on a regular basis, in order to work or benefit from the many services available in bigger cities. This could have increased the average mobility and hence, in the presence of asymptomatic or paucisymptomatic COVID-19 spreaders, increased overall contagion levels in the province.

The shares of population by age band also have some statistical significance in some of its specifications, with a negative and statistically significant sign for the 30–40, the 50–60 and the 60–70 (in 3.1 and 3.2) bands. This suggests that provinces with a bigger share of population among these age bands had fewer cases. This may be due to two factors. First, younger people are less likely to contract the severe form of the virus, and thus are less likely to be tested and reported in the statistics; this is especially relevant to the 30–40 band. Second, provinces with an older population may have had more casualties, and thus fewer cases given the higher fatality rate of the patients (and this is especially relevant to the 60–70 band age).

Finally, it is important to note that the R-squared in the different specifications also has a very interesting value, ranging from 0.707 for 3.4 to 0.743 for 3.1, suggesting that this operationalization of equation (1) is an appropriate model of the phenomenon.

In order to provide better estimates, we decided to replicate the analysis with another estimator. The literature has suggested that continuous zero to one variables (such as our dependent variable, which is expressed as a 0-1 bordered variable) may be better estimated through fractional response estimators. Papke and Wooldridge (1996) propose fractional probit models in a cross sectional context.

To improve the accuracy of the estimates, we run regressions through this estimator. Results of the marginal effects of these regressions are presented in Table 4, and show the same results once again, suggesting robustness in the findings. The total number of matches in January and February had a statistically very significant impact on the total cases in March and in March and April (4.1 and 4.3); once again the level of attendance does not have any statistically significant impact on this relationship (4.2 and 4.4).

All these pieces of evidence suggest that there was an important effect on the spread of COVID-19 due to the gatherings associated with football matches, while the total attendance of the same matches was not so important. In none of our specifications does the per capita monthly attendance of football events show a statistically significant impact on the COVID-19 cases per capita of the month after. While the coefficient of the variable computing total per capita attendance is always positive, and the low number of observations may lead to statistical significance lower than the thresholds usually adopted in the literature, this finding may also be attributed to the fact that at each match there are many people moving around and interacting with the fans, such as street food sellers, people in public transportation, gadget sellers, and so on, who may make the event of hosting of a match more relevant with regard to the spread of COVID-19 than the total number of people attending a match.

More precisely, the present analysis suggests with its preferred specification that for each extra match during the months of January and February in a province with 1,000,000 inhabitants, there were 234 more cases per 1,000,000 inhabitants in March, and 856.5 more in March and April. While we repeat that these estimates should be taken with caution, given the possible biases and issues discussed above, this seems nonetheless to be a significant impact, given that the average per capita infection rate in Italy during March was 0.02, and the minimum value registered in the same month was 0.001 (in the province of Isernia).

While our analysis relies on reported COVID-19 cases, a function of testing policy and capacity that varies greatly between countries and regions, it may be worthwhile to compare the effect we found with those of similar studies focused on other countries. Nonetheless, for the reasons we have already mentioned, we urge the reader to be cautious with such comparisons.

First of all, we should state that our estimate seems to be in line with results from Olczak et al. (2020) and Fischer (2021). Our coefficients may be considered as estimating an impact very close to the upper limit reported by the increase in the seven-day incidence between 3.6 and 6.4 in Germany (suggested in Fischer, 2021), and of 6 more cases per 100,000 inhabitants estimated in the UK (reported in Olczak et al., 2020). It is important to remember that our estimation of 234 and 856 more cases are referred respectively to the month of March and to March and April, while previous estimations are referred to a weekly increase. This means that, in both cases, assuming an average RT of 2.5 during the five weeks of March and of 1.4 during the following five weeks of April, our estimations are equivalent to estimating 6

Table 4. Fractional Probit Estimator: Determinants of Monthly COVID-19 Cases - Marginal Effects.

	(4.1) March Cases pc	(4.2) March Cases pc	(4.3) Mar-Apr Cases pc	(4.4) Mar-Apr Cases pc
ShPop 20-30	-0.827 (-1.18)	-1.030 (-1.40)	-2.460 (-0.90)	-3.029 (-1.06)
ShPop 30-40	-0.906*** (-2.86)	-0.934*** (-2.72)	-3.620*** (-3.20)	-3.610*** (-2.60)
ShPop 40-50	-0.0683 (-0.14)	-0.159 (-0.33)	-0.873 (-0.41)	-1.086 (-0.50)
ShPop 50-60	-0.987*** (-2.74)	-1.069*** (-2.91)	-3.957*** (-2.72)	-4.142*** (-2.78)
ShPop 60-70	-0.932* (-1.95)	-0.911* (-1.89)	-2.581 (-1.52)	-2.479 (-1.48)
ShPop over 70	-0.206 (-0.86)	-0.249 (-0.90)	-0.691 (-0.70)	-0.761 (-0.68)
PopDens	-0.00000316 (-1.23)	-0.00000330 (-1.25)	-0.0000176* (-1.84)	-0.0000177* (-1.81)
Income	0.00000213* (1.95)	0.00000173 (1.56)	0.0000143*** (3.60)	0.0000129*** (3.31)
Share of pop. in mun under 30k inhab	0.0243** (2.27)	0.0297*** (2.66)	0.0865** (2.15)	0.106** (2.44)
Feb Cases	0.0000690*** (4.95)	0.0000710*** (4.67)	0.000293*** (4.77)	0.000300*** (4.50)
Jan-Feb matches pc	234.0** (2.37)		856.5** (2.21)	
Jan-Feb public pc		0.0193 (0.68)		0.0595 (0.51)
Observations	98	98	98	98

Marginal effects; *t* statistics in parentheses (d) for discrete change of dummy variable from 0 to 1. Robust standard errors, observations weighted by province population size **p* < 0.1, ***p* < 0.05, ****p* < 0.01.

more cases in the week after a football match. While precise estimations of RT, given a testing campaign that was continually being fine-tuned, and thus constantly shifting, are difficult to derive, Riccardo et al. (2020) suggest that R0 varied between 2.5 in Tuscany and 3 in Lazio between 20 February and 24 March, with a rapid decline afterwards. If it is true that many Italian regions had a lower RT during this period, it is also important to highlight that since they had considerably fewer cases, these are weighted less in the total amount of cases. For all these reasons, our results seem to be comparable with the higher estimates already presented in the literature.

At the same time, when comparing these results, it is important to highlight two other points. First, Fischer's analysis focuses on the second German wave, a time in which COVID-19 was already well known, and people were more careful and taking many precautions. Our analysis, on the other hand, is focused on a period in which the presence of COVID was largely unknown or in any case neglected. Second, given the exponential nature of COVID infection rates, we expect such a multiplicative effect in the long term, even although it is hard to derive the exact extent of the exponent, also given changes in testing policy and variance among regions. Finally, it is also important to recognize that the degree of granularity of data used by Olczak et al. (2020) gives their analysis a specificity that is not achievable with data available for Italy during the same period.

Conclusions

Who knows whether we will manage to see
a stadium with 60,000 spectators again;
it would be positive in all senses.

- (In?)voluntary humor on an Italian football fan group

COVID-19 is a virus that is spread through the air, via droplets emitted by breathing, talking, chanting, and so on (Hossain & Faisal, 2021). For this reason, it is of the utmost importance to test the impact of mass events on the spread of COVID-19. Among such events, we consider football matches to be one of the most interesting cases, given their popularity in many European countries and their cultural importance for the population. In Europe, Italy seems to be the perfect case to test this relationship, given, on the one hand, Italians' tremendous interest in football and the absence of any other sport in the country with similar popularity, and, on the other, the dramatic spread of COVID-19 in the country during the first months of 2020.

In this study, using data about the attendance of football matches in Italy that occurred before the virus was given greater attention by the public and authorities, but after it was already present within Italian borders (La Rosa et al., 2020), we studied the impact of football events held in the different Italian provinces on the subsequent number of COVID-19 cases reported in the same provinces. Our best estimates, which also take into account the nature of the dependent variable, the total number of COVID-19 cases in February, the age distribution within provinces, population density, mean income, and the nature of the diffusion of the population within the province, show that for each football match played in January and February in a province there were over 234 more COVID-19 cases in March in the same province, and 856 over March and April. The same estimates suggest that the number of fans did not play a statistically significant role in this relationship, suggesting that hosting matches, with all that that entails, creates a spreading effect in the population over the subsequent months, rather than the actual attendance.

In other words, our evidence suggests that hosting matches, with its impact on mobility and congregation of people willing to trade or offer services to fans, creates the conditions for the spread of the virus, and over subsequent months (possibly those in which less attention was paid to the virus, above all) this opportunity for contagion creates multiple infections, significantly affecting the total rate of contagion in a province. In other words, using an argument similar to the one already expressed for re-opening schools (Alfano et al., 2021; Alfano, 2021), and suggested by Fischer (2021) with regard to stadiums in Germany, even if we make stadiums extremely secure, the real challenge would be to make the surrounding areas COVID-free, including all services used by the fans around the stadium, the public transportation needed to reach the venue, and so on. This implies that reducing the maximum number of people allowed in stadiums may be a sub-optimal strategy for fighting the spread of COVID-19.

Overall, all these pieces of evidence suggest that, as happened in England (Olczak et al., 2020), in Germany (Fischer, 2021), and in the USA (although in this case the literature, namely the studies by Ahammer et al., 2020 and Cardazzi et al., 2020, focused on indoor events), in Italy too, football events had a significant impact on the spread of COVID-19.

The initial implication of this is straightforward: opening stadiums is a decision that governments and local authorities must approach with caution, to avoid wasting many sacrifices already made in the fight with the virus. As noted by Olczak et al. (2020), while the decision to keep stadiums closed undoubtedly has significant economic consequences, especially for sports leagues and clubs, which rely on attendance of matches to survive financially, the estimated impact of these mass events on the spread of the virus is too important to be overlooked.

This study has some possible limitations, which it is important to warn the reader about. First, while football is widely popular in Italy, as highlighted by previous literature, and there is no other sport that may really be compared to it in terms of interest (Boeri & Severgnini, 2014; ISTAT, 2005), it is possible that in Italian provinces without a competitive football team there was extra attendance of other sporting events that contributed to spreading the virus. This would of course introduce a bias in the present analysis. While, as said, there are reasons to believe that no sports event in Italy has an attendance even remotely comparable to that of football, we believe it is worth warning the reader about this potential issue.

Second, it has been suggested that COVID data, especially those regarding the beginning of the first wave, are very unreliable, due to the limited testing and the policies in place at the time. Unfortunately, while we recognize that this is an important limit of the analysis that the reader should be made aware of, we also believe that there is not much that can be done to solve this issue, and it is also worth recalling that even if it is likely that COVID-19 infections were underestimated in early 2020, there are no reasons to believe that the actual figures are not highly correlated with those reported, ensuring our estimates maintain a certain significance.

Follow-up studies may focus on other national cases, trying to expand our findings, and comparing our estimates with those regarding another case, or focusing on other kinds of outdoor or indoor mass events, such as concerts.

Author Note

Author wish to thank Mariana Gabriela Cesario for tireless help in constructing the dataset used in this work and two anonymous referees for their valuable help in improving the present work. All errors remains of the author.


Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Vincenzo Alfano  <https://orcid.org/0000-0002-4981-748X>

Notes

1. With this word in the present work we refer to the sport Americans call soccer.
2. <https://www.wsj.com/articles/the-soccer-match-that-kicked-off-italys-coronavirus-disaster-11585752012> (URL accessed on 19/6/2021). Other articles in the press include ESPN (<https://www.espn.co.uk/football/atalanta/story/4318221/one-year-after-game-zero-bergamo-finds-joy-in-atalanta> URL accessed on 21/6/2021) and *La Repubblica* (https://milano.repubblica.it/cronaca/2020/03/20/news/atalanta-valencia_a_san_siro_il_detonatore_del_contagio-300898801/ URL accessed on 21/6/2021).
3. Please also notice that, unlike Olczak et al. (2020), we did not include in the analysis away games. This choice is dictated by the fact that anecdotal evidence suggests that, aside from the top 2-3 teams in Serie A, teams do not attract spectators away. This is also confirmed by a study from Stadium Guide published on <https://www.stadiumguide.com/these-clubs-pull-in-the-crowds-away-from-home/> (url visited on 23/09/2021).
4. These provinces are Valle d'Aosta, Asti, Biella, Caltanissetta, Gorizia, Oristano, Pavia, Ragusa and Sud Sardegna. As a robustness check we replicate the regression analysis including these provinces in the data, obtaining very similar results, which have not been included for reasons of space, but are available upon request.
5. Please note that the share of citizens under 20 years old is not included, in order to avoid perfect multicollinearity.

References

- Ahammer, A., Halla, M., & Lackner, M. (2020). *Mass gatherings contributed to early COVID-19 spread: Evidence from US Sports*. Covid Economics, CEPR. <https://voxeu.org/article/mass-gatherings-contributed-early-covid-19-mortality>
- Alfano, V., Cicatiello, L., & Ercolano, S. (2021). School openings and the COVID-19 outbreak in Italy. A provincial-level analysis using the synthetic control method. *Health Policy*, 125(9), 1200–1207. Published online 2021 Jul 2. <https://doi.org/10.1016/j.healthpol.2021.06.010>
- Alfano, V., & Ercolano, S. (2020). The efficacy of lockdown against COVID-19: A cross-country panel analysis. *Applied Health Economics and Health Policy*, 18, 509–517. <https://doi.org/10.1007/s40258-020-00596-3>
- Alfano, V. (2021). The effects of school closures on COVID-19: A cross-country panel analysis. *Appl Health Econ Health Policy*, forthcoming. <https://doi.org/10.1007/s40258-021-00702-z>
- Alfano, V., & Ercolano, S. (2021a). Stay at home! governance quality and effectiveness of lockdown. *Social Indicator Research*, 1–23. <https://doi.org/10.1007/s11205-021-02742->
- Alfano, V., & Ercolano, S. (2021b). Social capital, quality of institutions and lockdown. Evidence from Italian provinces. *Structural Change and Economic Dynamics*, 59, 31–41. <https://doi.org/10.1016/j.strueco.2021.08.001>
- Boeri, T., & Severgnini, B. (2014). The decline of professional football in Italy. In J. Goddard, & P. Sloane (Ed.), *Handbook on the economics of professional football* (chap. 19, pp. 322–335). Edward Elgar Publishing.
- Brown, O. R. (2020). Asymptomatic COVID-19; We don't know what we don't know. Preprints, 2020070681. <https://doi.org/10.20944/preprints202007.0681.v1>
- Cardazzi, A., Humphreys, B. R., Ruseski, J. E., Soebbing, B., & Watanabe, N. S. (2020). Professional sporting events increase seasonal influenza mortality in US cities (Technical Report 3628649). SSRN, 2020.
- Cruz, B. S., & de Dias, O. M. (2020). COVID-19: From outbreak to pandemic. *GSI*, 8(3), 2230–2238. <https://doi.org/10.11216/gsj.2020.03>
- Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., CMMID COVID-19 working group and Eggo, R. M. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26, 1205–1211. <https://doi.org/10.1038/s41591-020-0962-9>
- Elgar, F. J., Stefaniak, A., & Wohl, M. J. A. (2020). The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries. *Social Science & Medicine*, 263. <https://doi.org/10.1016/j.socscimed.2020.113365>
- Ercolano, S. (2012). Italian dualism in foreign scholars' analyses. *Rivista Economica del Mezzogiorno*, 3(3), 411–444. <https://www.rivisteweb.it/doi/10.1432/38970>
- Figueiredo, A. M., Daponte Codina, A., Figueiredo, M., Saez, M., & Cabrera León, A. (2020). Impact of lockdown on COVID-19 incidence and mortality in China: An interrupted time series study. *Bulletin of the World Health Organization*. E-pub: 6 April. <https://doi.org/10.2471/BLT.20.256701>
- Fischer, K. (2021). *Thinning out spectators: Did football matches contribute to the second COVID-19 wave in Germany?* <https://ssrn.com/abstract=3793379>

- Han, P., & Ivanovski, S. (2020). Saliva—friend and foe in the COVID-19 outbreak. *Diagnostics, 10*(5), 290. <https://doi.org/10.3390/diagnostics10050290>
- Hossain, M., & Faisal, N. H. (2021). Modeling aerosol cloud aerodynamic during human coughing, talking, and breathing actions. *AIP Advances, 11*(4). <https://doi.org/10.1063/5.0042952>
- ISTAT (2005). *Lo sport che cambia*. Istituto nazionale di Statistica.
- ISTAT and Italian Ministry of Health. (2020). Primi risultati dell'indagine di sieroprevalenza sul SARS-COV-2 (Technical report).
- La Rosa, G., Mancini, P., Bonanno Ferraro, G., Veneri, C., Iaconelli, M., Bonadonna, L., Lucentini, L., & Suffredini, E. (2020). SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. *Science of the Total Environment, 750*, 141711. <https://doi.org/10.1016/j.scitotenv.2020.141711>
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of Travel Medicine, 27*(3). <https://doi.org/10.1093/jtm/taaa037>.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine, 172*(9), 577–582. <https://doi.org/10.7326/M20-0504>
- Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 97*(4), 558–625. <https://doi.org/10.2307/2342192>
- Nicastri, E., D'Abramo, A., Faggioni, G., De Santis, R., Mariano, A., Lepore, L., Molinari, F., Petralito, G., Fillo, S., Munzi, D., Corpolongo, A., Bordi, L., Carletti, F., Castiletti, C., Colavita, F., Lalle, E., Bevilacqua, N., Giancola, M. L., Scorzolini, L.,..., On Behalf Of Inmi And The Italian Army Covid-Study Groups. (2020). Coronavirus disease (COVID-19) in a paucisymptomatic patient: Epidemiological and clinical challenge in settings with limited community transmission, Italy. *Eurosurveillance, 25*, 2000230. <https://doi.org/10.2807/1560-7917.ES.2020.25.11.2000230>
- Olczak, M., Reade, J., & Yeo, M. (2020). *Mass outdoor events and the spread of an airborne virus: English football and Covid-19*. Covid Economics, CEPR. <https://voxeu.org/article/spread-covid-19-and-attending-football-matches-england>
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*, 619–632. [https://doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<619::AID-JAE418>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1)
- Reade, J. J., Schreyer, D., & Singleton, C. (2020). Stadium attendance demand during the COVID-19 crisis: Early empirical evidence from Belarus. *Applied Economics Letters, 28*(18), 1542–1547. <https://doi.org/10.1080/13504851.2020.1830933>
- Reade, J. J., & Singleton, C. (2020). Demand for public events in the COVID-19 pandemic: A case study of European football. *European Sport Management Quarterly, 21*(3), 391–405. <https://doi.org/10.1080/16184742.2020.1841261>

- Riccardo, F., Ajelli, M., Andrianou, X., Bella, A., Del Manso, M., Fabiani, M., Bellino, S., Boros, S., Urdiales, A. M., Marziano, V., Rota, M. C., Filia, A., D'Ancona, F., Siddu, A., Punzo, O., Trentini, F., Guzzetta, G., Poletti, P., Stefanelli, P., ..., COVID-19 working group. (2020). *Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic*. <https://www.medrxiv.org/content/10.1101/2020.04.08.20056861v1>
- Sassano, M., McKee, M., Ricciardi, W., & Boccia, S. (2020). Transmission of SARS-CoV-2 and other infections at large sports gatherings: A surprising gap in our knowledge. *Frontiers in Medicine*, 7. <https://doi.org/10.3389/fmed.2020.00277>
- Stoecker, C., Sanders, N. J., & Barreca, A. (2016). Success is something to sneeze at: Influenza mortality in cities that participate in the super bowl. *American Journal of Health Economics*, 2(1), 125–143. https://doi.org/10.1162/AJHE_a_00036