**BMJ Open**

**Effect of the Tokyo 2020 Summer Olympic Games on COVID-19 incidence in Japan: a synthetic control approach**

Daisuke Yoneoka, Akifumi Eguchi, Kentato Fukumoto, Takayuki Kawashima, Yuta Tanoue, Takahiro Tabuchi, Hiroaki Miyata, Cyrus Ghaznavi, Kenji Shibuya, Shuhei Nomura

**ABSTRACT**

**Background** The Tokyo 2020 Summer Olympic Games (23 July–8 August 2021) were held in the middle of Japan’s fifth wave of COVID-19, when the number of cases was on the rise, and coincided with the fourth state of emergency implemented by the host city, Tokyo.

**Aim** This study aimed to assess whether the hosting of the Games was associated with a change in the number of COVID-19 cases in Japan using a synthetic control method.

**Methods** A weighted average of control countries with a variety of predictors was used to estimate the counterfactual trajectory of daily COVID-19 cases per 1 000 000 population in the absence of the Games in Japan. Outcome and predictor data were extracted using official and open sources spanning several countries. The predictors comprise the most recent country-level annual or daily data accessible during the Games, including the stringency of the government’s COVID-19 response, testing capacity and vaccination capacity; human mobility index; electoral democracy index and demographic, socioeconomic, health and weather information. After excluding countries with missing data, 42 countries were ultimately used as control countries.

**Results** The number of observed cases per 1 000 000 population on the last day of the Games was 109.2 (7-day average), which was 115.7% higher than the counterfactual trajectory comprising 51.0 confirmed cases per 1 000 000 population. During the Olympic period (since 23 July), the observed cumulative number of cases was 61.0% higher than the counterfactual trajectory, comprising 143 072 and 89 210 confirmed cases (p=0.023), respectively. The counterfactual trajectory lagged 10 days behind the observed trends.

**Conclusions** Given the increasing likelihood that new emerging infectious diseases will be reported in the future, we believe that the results of this study should serve as a sentinel warning for upcoming mega-events during COVID-19 and future pandemics.

**INTRODUCTION**

The COVID-19 pandemic poses major political, socioeconomic, scientific and public health challenges to countries across the globe. Large-scale mass gathering events, such as sporting, musical and religious functions, have historically been a major source of infectious disease transmission and represent a major public health challenge.

The Tokyo 2020 Summer Olympic Games (23 July–8 August 2021) were held in the middle of Japan’s fifth wave of COVID-19, when case numbers were on the rise, and coincided with the fourth state of emergency implemented by the host city, Tokyo. The Japanese public, government and scientific community were split in terms of whether or not they believe the Olympics should be hosted because of COVID-19 infection concerns. The government went to great lengths to minimise risk during the event: active vaccination was recommended to all those affiliated with the Games, including the athletes, and all possible countermeasures against infection were taken, including the banning of spectators. The evidence base concerning risk assessment and decision-making to minimise the transmission of infectious diseases during mega-events is still evolving and needs to be expanded.
For example, a series of studies concerning COVID-19 infection risk in upcoming sporting mega-events such as FIFA World Cup Qatar 2022 conducted by Dergaa and colleagues concluded that stringent public health policies such as a tight ‘bubble system’ for players were key components to ensuring the successful containment of COVID-19. With the ongoing pandemic and the inevitable future spread of emerging infections around the world, it is important to evaluate the impact of the Tokyo Games on the spread of COVID-19.

We examined the association between the hosting of the Tokyo Games and the daily number of COVID-19 cases in Japan by using a synthetic control method (SCM) to approximate the counterfactual trend in the daily number of COVID-19 cases, assuming the absence of the Games. The SCM has been widely used in the social sciences and is increasingly used in epidemiology to assess the impact of public health interventions such as tobacco control policies, soft drink taxation, social welfare reform and COVID-19 interventions.

According to the Japanese Olympic Committee, the total number of infected people involved in the Olympics was 547.14 However, rather than transmission from infected Olympic affiliates to Japanese locals, we hypothesise that the hosting of the Tokyo Games may have influenced the behavioural psychology of the public, which had previously been practising self-restraint under the state of emergency. If the Games indirectly encouraged lower compliance with public health guidance, there may have been a tangible impact on Japan’s infection landscape.

METHODS

Approach

In this study, we employed the SCM, which is particularly suitable for population-level studies and allows for appropriate comparisons when random assignment of an intervention is not possible. While other modelling methods such as the difference-in-differences approach are available, the SCM can be easily applied to specific cases where there are multiple control units and it is difficult to select the optimal group of controls for comparison with only one treated unit.

We evaluated the impact of the Tokyo Games on daily confirmed cases of COVID-19 (per 1 000 000) by estimating the counterfactual trajectory of cases assuming the Olympics were not held in Japan and comparing it with the observed trajectory. Daily confirmed cases of COVID-19 were smoothed by using the 7-day moving average in order to adjust for day-of-the-week effects. The counterfactual (or synthetic control) consists of a weighted combination of the COVID-19 incidence in countries in which the Games were not held (hereafter, ‘donor pool’). If the synthetic control closely reproduces the observed trajectory during the period before the start of the Olympics (preintervention period) in Japan, we can have confidence in the validity of the estimated counterfactual trajectory after the start of the Tokyo Games. An important advantage of this approach is that it does not require the identification of an individual comparison country that is sufficiently similar to the Olympics host (ie, Japan).

After removing countries that had ≥10% missing values in the number of COVID-19 tests, vaccination rate and stringency index or countries that have any completely-missing outcomes, the following 42 countries were included in the donor pool: Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Czech Republic, Dominican Republic, Ecuador, Estonia, France, Germany, Greece, Guatemala, Hong Kong, Indonesia, Ireland, Israel, Italy, Latvia, Lithuania, Malaysia, Malta, Mexico, New Zealand, Norway, Paraguay, Peru, Russia, Slovakia, Slovenia, South Korea, Switzerland, Thailand, Trinidad and Tobago, Turkey, Ukraine, United Kingdom, United States, Uruguay and Vietnam. The candidate predictor set and further details can be found in the online supplemental appendix.

We constructed the counterfactual as follows: first, we used the donor pool and the candidate predictor set to estimate the weights that are optimised to attain the smallest mean squared prediction error (MSPE) during the preintervention period. More precisely, the SCM completes the following two-step minimisation task iteratively to find optimal values of the predictor weight vector $V \in \mathbb{R}^M$ and the country weight vector $W \in \mathbb{R}^J$:

$$w^* (V) = \min_{W} \sum_{j=1}^{M} \sum_{t=2}^{T} w_j (Y_{jt} - \sum_{m=1}^{M} \mathbf{w}_m X_{mt})^2,$$

where $w^* (V) = (w_2^*, \ldots, w_{J+1}^*)^T$, $V = (v_1, \ldots, v_M)^T$, $M$ is the number of predictors, $J$ is the number of control countries and $X_i$ and $X_J$ are the predictor vectors for the treated country (ie, Japan) and the other $J$ countries, respectively. Then, the weight vector $V$ is optimised via the following:

$$\min_{V} \sum_{j=1}^{J} \sum_{t=2}^{T} (Y_{jt} - \sum_{m=1}^{M} \mathbf{w}_m X_{mt})^2,$$

where $t \in \{1, \ldots, T_0\}$ is the preintervention period and $Y_{1J}$ and $Y_{Jt}$ are the number of COVID-19 cases in Japan and the other countries at time $t$. The conjugate gradient method was used for the optimisation. Kim10 proved that the SCM proposed by Abadie et al18 is equivalent with the (restricted version of the) L1 penalty. This implies that the method tends to assign positive weights to only a subset of the donor pool and predictors; thus, it works similarly to Lasso-based (Lasso: least absolute shrinkage and selection operator) variable selection. Second, by using the country weights (online supplemental figure 1 and online supplemental table 2), we estimate the path of the counterfactual in the period after the start of the Games (postintervention period). To check whether there is a meaningful effect due to the intervention, as shown by Abadie et al18, Fisher’s exact $p$ values were calculated by dividing the postintervention MSPE by the preintervention MSPE. For inference, the SCM was repeated by regarding one country in the donor pool as
a treated country and the remaining countries and Japan as an alternative donor pool generating placebo synthetic controls: online supplemental figure. *R* V.4.0.5 and the *tidysynth* package were used for the analysis.

If the effect of the Games on daily confirmed cases is mediated by behavioural factors, these changes may not necessarily manifest on the opening day of the Tokyo Games; for example, behaviours could change after the announcement that it would be held without spectators (8 July). Furthermore, there is a time lag—between infection, onset of disease and testing—before the effects of the Games manifest in the number of daily confirmed cases. Therefore, we performed a sensitivity analysis in which we also constructed counterfactuals in the same manner for the 7 days before and after the opening ceremonies as the intervention timing.

**Data**

In this study, outcome and predictor data were extracted using official and open sources spanning dozens of countries. Data on the daily confirmed cases of COVID-19, the stringency of the government’s COVID-19 response, testing capacity and vaccination capacity were obtained from Our World in Data; data on human mobility was obtained from Google’s COVID-19 Community Mobility Reports; other demographic, socioeconomic, health and weather data were obtained from Our World in Data, the World Bank and the United Nations. The electoral democracy index, a measure of political freedom, was downloaded from the Varieties of Democracy (V-Dem) project. The online supplemental appendix presents summary statistics for all predictors.

**Patient and public involvement**

No patients or members of the public were involved in the design, analysis or dissemination plan of our study. Our study relies on population-level data from a large number of countries, and all data used were obtained from existing public sources.

**RESULTS**

Figure 1A shows the observed and counterfactual (ie, the weighted average of 42 countries) trajectories of daily cases per 1000000 on the vertical axis, where the days relative to Olympic beginning ceremony are shown on the horizontal axis and a vertical line is drawn on day 0. The counterfactual trajectory tracks along the observed trajectory well prior to the Games (left side of the vertical line). Figure 1B displays the differential between the observed and counterfactual trajectories over the same period. The difference between the observed and counterfactual values became positive after the start of the Olympics (right side of the vertical line). The SCM revealed that as of the closing day of the Olympics (day 16), the number of daily observed cases per 1000000 population was 109.2, which was 115.7% higher than the counterfactual trajectory comprising 51.0 confirmed cases per 1000000 population. During the Olympic period, the observed cumulative number was 60.4% higher than the counterfactual scenario, with 143072 and 89210 confirmed cases (p=0.023), respectively. There was a lag of approximately 10 days between the observed and counterfactual confirmed cases: the counterfactual daily total would have been approximately 6400 on the day of the closing ceremonies, but in reality, it was reached 10 days earlier (29 July 2021).

The results of the sensitivity analyses (online supplemental appendix), in which counterfactuals were constructed by shifting the timing of the intervention (ie, the opening ceremonies) by ±7 days, were consistent with our main findings.

**DISCUSSION**

Our intention was to add to the evidence base concerning risk assessment and decision-making to minimise the transmission of infectious diseases during mega-events. Using the SCM, we assessed whether the hosting of the Tokyo 2020 Olympic Games was associated with an increase in the number of COVID-19 cases in Japan. Based on the
difference between the observed and counterfactual trajectories, we found that approximately 53,900 excess infections were reported over the course of the Olympic period. This estimate was consistent with prior results reported by Yamamoto et al., which compared Tokyo with other prefectures in Japan and found that the number of observed cases per 1,000,000 population in Tokyo was approximately 100, while the number of counterfactual cases per 1,000,000 population was approximately 50, as of the closing day of the Olympics.

Though previous research has found that the Olympics and major sporting events do not substantially increase the risk of infectious disease outbreaks, never before have the Games been hosted during a pandemic of this scale. In Japan, the Olympic ‘bubble’ was partitioned from society and was largely successful in insulating new cases at the national level. In the wake of the more infectious Delta variant, insufficiently preintervention infection rates and predictors of the donor pool in this study, of which five countries (Germany, Hong Kong, Italy, Thailand, and South Korea) received negligible weights. In each of these countries, the daily infections were on a slight upward trend during the Olympics, including with high-profile reporting of Olympians violating infection control rules, may have had an impact on adherence in the general public. Finally, despite the lack of spectators and tourists in Tokyo during the Games, the human mobility during the Olympics was greater than during the previous three states of emergency.

The exceptions made for athletes to enter Japan when most non-Japanese persons were barred entry, in conjunction with high-profile reporting of Olympians violating infection control rules, may have had an impact on adherence in the general public. Finally, despite the lack of spectators and tourists in Tokyo during the Games, the human mobility during the Olympics was greater than during the previous three states of emergency. In the midst of the rise of the more infectious Delta variant, insufficiently reduced levels of domestic mixing would directly increase the risk of exposure to COVID-19 among locals.

A previous modelling study of the Games has found that infection prevention measures were ostensibly responsible for controlling transmission at the national level, but the surge in Tokyo case numbers was likely related to local transmission, which may have been attributable to mobility and spread of the Delta variant. Notably, as of July 2021, Japan’s vaccine rollout was relatively slow, and major efforts were made to get as many locals vaccinated as possible before the start of the Games. At the time of the opening ceremony, the government achieved a two-dose immunisation rate of only approximately 30% among those living in Japan. In order to minimise the spread and burden of infectious diseases during mega-events, efforts should be made prior to the events to improve vaccination coverage in the host areas and the broader population, beyond the recommendation of the vaccine of athletes and other relevant persons including staff. In addition, we encourage consideration of additional health needs of not only those participating in the events but also those who work and reside around the event locations as well as the communities where most human traffic occurs as a result of the events. Preparations should be made at the clinic and hospital levels to accommodate an influx of patients who may be affected by changes in social behaviour that can precipitate or further propagate disease outbreaks.

Limitation
We note that it was challenging for the SCM to rule out the impact of an event that may have occurred almost simultaneously with the Games. For instance, several other events such as consecutive holidays occurring between 22 and 25 July 2021 and the beginning of summer vacation may have confounded our findings. To address this concern, we implemented two more SCM analyses where we shifted the timing of the intervention from 23 July to 16 and 30, respectively. The results did not change our interpretation of the results (online supplemental figures 3 and 4). Second, we were unable to make definitive assessments of individual-level mechanisms regarding the relationship between the Games and COVID-19 cases. Our analyses did not include data regarding behavioural changes or routes of transmission during the Olympics. Further research to address these questions is merited. Finally, 42 countries were included in the donor pool in this study, of which five countries (Germany, Hong Kong, Italy, Thailand and South Korea) received negligible weights. In each of these countries, the daily infection cases were on a slight upward trend during the Olympics, with the exception of Thailand, which was experiencing a sharp increase in case load. Hong Kong had a small number of daily cases (less than 10). The weights were determined by the list of predictors, including the lagged infection status, to attain the optimal MSPE during the pre-Olympic period. Note that the SCM estimates weights such that the weighted averages of the preintervention infection rates and predictors of the donor pool were close to those of Japan. The weights are estimated using covariates included in the predictor set, but the number of covariates we included in the estimation process is limited, and we cannot account for unobserved covariates that may have played a role in the transmission of COVID-19. However, based on the predictors we considered,
the current result was optimal in the sense that the SCM estimated country-specific weights that minimised the pre-threshold MSPE.

Conclusion
Using a synthetic control design, we assessed whether the hosting of the Games was associated with a change in the threshold MSPE.

Author affiliations
1Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo, Japan
2The Tokyo Foundation for Policy Research, Tokyo, Japan
3Department of Sustainable Health Science, Chiba University, Chiba, Japan
4Department of Political Studies, Gakushuin University, Tokyo, Japan
5Department of Mathematical and Computing Science, Tokyo Institute of Technology, Tokyo, Japan
6Institute for Business and Finance, Waseda University, Tokyo, Japan
7Department of cancer epidemiology, Osaka International Cancer Institute, Osaka, Japan
8Department of Health Policy and Management, Keio University, Tokyo, Japan
9Medical Education Program, Washington University in St Louis School of Medicine, St Louis, Missouri, USA

Twitter Takahiro Tabuchi @TakahiroTabuchi

Contributors
DY, AE, KE, TK, YT, TT, HM, KS and SN conceived and designed the study and take responsibility for the accuracy of the data analysis. DY, AE, KE, TK, YT, CG and SN analysed and interpreted the data. DY and AE conducted statistical analysis and drafted the article. DY, AE, KE, TK, YT, TT, HM, CG, KS and SN made critical revision of the manuscript for important intellectual content and gave final approval for the manuscript. SN is responsible for the overall content as guarantor.

Funding
The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests
None declared.

Patient and public involvement
Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication
Not applicable.

Ethics approval
Not applicable.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
Data are available in a public, open access repository. We have uploaded the dataset and associated R programs in the author’s open github account (https://github.com/kingwert/R/tree/master/Synthetic_Olympic).

Supplemental material
This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

ORCID iDs
Daisuke Yoneoka http://orcid.org/0000-0002-3525-5092
Kantaro Fukumoto http://orcid.org/0000-0003-3704-9054
Shuhei Nomura http://orcid.org/0000-0002-2963-7297

REFERENCES
19 Kinn D. Synthetic control methods and big data. ArXiv 2018;1803.0096v1.
33 The Mainichi Shimbun. Outdoor drinking party in athletes village is investigated. Tokyo: The Mainichi Shimbun, 2021. https://mainichi.jp/english/articles/20210801/p2g/00m/0sp/046000c