



RESEARCH ARTICLE

REVISED **Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England [version 2; peer review: 3 approved]**

Robert Smith ¹, Paul Schneider ¹, Alice Bullas ², Steve Haake ², Helen Quirk ², Rami Cosulich¹, Elizabeth Goyder ¹

¹School of Health and Related Research, University of Sheffield, Regents Court, Sheffield, S1 4DA, UK

²Advanced Wellbeing Research Centre, Sheffield Hallam University, Olympic Legacy Park, Sheffield, S9 3TU, UK

v2 **First published:** 16 Jan 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.1>
Latest published: 18 Jun 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.2>

Abstract

Background: parkrun has been successful in encouraging people in England to participate in their weekly 5km running and walking events. However, there is substantial heterogeneity in parkrun participation across different communities in England: after controlling for travel distances, deprived communities have significantly lower participation rates.

Methods: This paper expands on previous findings by investigating disparities in parkrun participation by ethnic density. We combined geo-spatial data available through the Office for National Statistics with participation data provided by parkrun, and fitted multivariable Poisson regression models to study the effect of ethnic density on participation rates at the Lower layer Super Output Level.

Results: We find that areas with higher ethnic density have lower participation rates. This effect is independent of deprivation.

Conclusions: An opportunity exists for parkrun to engage with these communities and reduce potential barriers to participation.

Keywords

parkrun, Physical Activity, Ethnic Density, Deprivation

Open Peer Review

Reviewer Status

	Invited Reviewers		
	1	2	3
version 2 (revision) 18 Jun 2020	 report	 report	 report
version 1 16 Jan 2020	 report	 report	 report

- Anne Grunseit** , The Australian Prevention Partnership Centre, NSW, Australia
University of Sydney, Sydney, Australia
- Gozde Ozakinci** , University of St Andrews,
St Andrews, UK
- Stephen Senn** , Consultant Statistician,
Edinburgh, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert Smith (rasmith3@sheffield.ac.uk)

Author roles: **Smith R:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Schneider P:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bullas A:** Project Administration, Resources; **Haake S:** Project Administration, Resources, Writing – Review & Editing; **Quirk H:** Project Administration, Resources, Writing – Review & Editing; **Cosulich R:** Writing – Review & Editing; **Goyder E:** Supervision, Writing – Review & Editing

Competing interests: R.S. and P.S. have no competing interests. S.H. is chair, A.B. and H.Q. are deputy chairs, and L.G. is a member of the parkrun research board.

Grant information: R.S., P.S. & R.C. are joint funded by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science [108903] and the University of Sheffield.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Smith R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Smith R, Schneider P, Bullas A *et al.* **Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England [version 2; peer review: 3 approved]** Wellcome Open Research 2020, 5:9 <https://doi.org/10.12688/wellcomeopenres.15657.2>

First published: 16 Jan 2020, 5:9 <https://doi.org/10.12688/wellcomeopenres.15657.1>

REVISED Amendments from Version 1

We have made minor changes to the text to provide greater clarity as per the very helpful comments provided by the reviewers. There are no changes to the main findings, data or methods used.

Any further responses from the reviewers can be found at the end of the article

Introduction

parkrun is a collection of free mass participation 5km running events that takes place every Saturday morning. There are currently over 500 locations in England, with a combined weekly attendance of over 100,000. parkrun has been identified as being successful at engaging with individuals who may not otherwise have taken part in organised physical activity^{1,2}, and there is some evidence that it has increased overall physical activity levels in participants³. Overall, there is a consensus that parkrun has huge public health potential⁴.

However, qualitative research in Sheffield⁵ and other areas of the United Kingdom⁶ identified that parkruns located in more deprived areas have lower attendances, and that ethnic diversity in parkrun was limited. This leads to concern that as with many public health interventions, parkrun is “likely to be responsible for significant intervention generated inequalities in uptake of opportunities for physically active recreation”⁵.

Undertaking quantitative analysis of the determinants of participation in parkrun is therefore long overdue. Apart from a single previous study from Australia⁷, with substantial limitations including, as noted by the authors, that “The sample was limited to a non-random sample of parkrun participants in one State of Australia and may not be generalizable to other parkrun populations.” (p.21), no other studies have attempted to identify the determinants of participation in parkrun.

Our previous work revealed that there is substantial heterogeneity in parkrun participation across different communities

in England: after controlling for geographical distance to nearest event, deprived communities have significantly lower participation rates⁸. The analysis was able to quantify, for the first time, how participation in parkrun varied in different communities in England. However, the analysis only explored the relationship between participation, access and deprivation and did not consider ethnic density as a potential determinant of participation in parkrun. Evidence from survey data shows that non-White-British individuals in England are less likely to be physically active, and to engage in sport in general⁹. We thus hypothesised that at the community level, areas with higher ethnic density have lower levels of participation in parkrun.

Methods

Ethical statement

Ethical approval was obtained from the Sheffield Hallam University Ethics Committee (ER10776545). We did not collect any personal information, but only used aggregate secondary data. The parkrun Research Board approved this research project, and three of its members (AMB; EG, SSSJH) were actively involved in it.

Data sources

We undertook an ecological analysis of parkrun participation in England in 2018. Data was obtained from multiple sources (see [Table 1](#)) for the 32,844 Lower layer Super Output Areas (LSOAs) in England, each of which is a geographical area containing around 1,500 people. parkrunUK provided data on the number of parkrun finishers from each LSOA in England between the 1st January and 10th December 2018, which we use as a proxy for parkrun participation, although we appreciate that people participate in parkrun in other ways (e.g. volunteering). We also used parkrun event location data, which are publicly available on the parkrunUK website.

The rest of the data, including Index of Multiple Deprivation (IMD) Score, Ethnic Density, Rural-Urban Classification, Population Density, Percentage Working Age and LSOA centroids were obtained from the Office of National Statistics (ONS). Descriptions of variables and sources are listed in [Table 1](#),

Table 1. Variables used in the analysis.

Variable	Description	Source
Finishers	Number of parkrun finishers during period	parkrunUK (2018)
IMD score	Index of Multiple Deprivation score	ONS (2019)
Population	Total number of inhabitants	ONS (2019)
Pop density	Population density (pop/km ²)	ONS (2019)
Rural-urban classification	Rural-urban classification (binary)	ONS (2019)
Ethnic density	Proxy: Percentage of population non-White-British	ONS (2019)
Distance	Distance from LSOA centroid to nearest parkrun	derived
Non-working-age	Percent of population not 16-65	ONS (2019)
Participation rate	Number of finishes/1000 population	derived

IMD, Index of Multiple Deprivation; LSOA, Lower layer Super Output Area; ONS, Office for National Statistics.

and all data is provided open source as *Underlying data* and on the author's GitHub page (<https://github.com/bitowaqr/DoPE>)¹⁰.

Data analysis

The merged data-set contains complete data for all LSOAs, and therefore all LSOA were included within the analysis, which was conducted using R software environment version 3.5.1 (2018-07-02)¹¹. We first used a simple colour plot to display the relationship between deprivation, ethnic density and parkrun participation graphically using ggplot¹². We then used Poisson regression models, commonly used when working with count data, to estimate the relationship between ethnic density, deprivation and parkrun participation, controlling for potential confounding variables including: population density, population, age and distance to nearest parkrun event.

Results

Descriptive statistics

Descriptive statistics are shown in [Table 2](#). Participation in parkrun varies across LSOAs, with around half of all communities (LSOA) averaging less than one finisher per week per 1,000 people. Approximately a quarter average between one and two finishers, and around an eighth between two and three finishers. There is considerable variation in ethnic density, with most LSOAs having a large majority of White-British residents, and few areas having over 50% non-White-British residents. Deprivation score is positively skewed, meaning that most areas have low deprivation, with a few very deprived areas. Finally, around 70% of LSOAs are within 5km, the parkrun distance, of a parkrun. Again, this is positively skewed with half of all LSAOs being within 3.5km of their nearest event.

There is a negative correlation between participation and the following: deprivation (IMD), distance to nearest parkrun, population density and ethnic density. Ethnic density is strongly positively correlated with population density, negatively correlated with percentage non working age, and moderately positively correlated with IMD, suggesting that areas with higher ethnic density are more densely populated overall, more deprived and have a higher percentage of working age people.

The colour plots in [Figure 1](#) show the participation rates for LSOA by deprivation and ethnic density for urban and rural areas¹³. Yellow, green and blue indicate high, moderate and low levels of participation respectively. The plot shows that participation is generally greatest in areas that have low levels of deprivation and low levels of ethnic density (bottom left), and lowest in areas with high levels of deprivation and high ethnic density (top-right). Areas with either high deprivation, or high ethnic density, tended to have low participation, suggesting that both are important independently. The relationship was robust to urban major areas and urban minor areas but did not hold in rural areas where data was more limited. It is important to note that we do not control for other factors, such as the age of residents or the population density, which are known confounders of this relationship.

Poisson model

The results of three Poisson regression models are shown in [Table 3](#). All models include the control variables: population density, distance to nearest event and percentage of the population of non-working age. Model 1 includes IMD Score, Model 2 includes ethnic density and Model 3 includes both IMD and ethnic density. All coefficients are significant at the $p < 0.01$ level.

Model 1 shows that, controlling for population density, distance to nearest event and age of population, areas with higher IMD (more deprived) have lower participation.

Model 2 shows that, with the same controls, areas with higher ethnic density have lower participation.

Model 3 shows that when both independent variables (IMD and ethnic density) are included their coefficients decrease, suggesting that some of the effect previously attributed to deprivation is indeed due to lower participation in areas with higher ethnic density.

Discussion

Our findings show that more deprived areas and areas with higher ethnic density have lower participation rates. This effect

Table 2. Descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Finishers	32,844	123.6	128.9	0	33	86	172	1,659
IMD score	32,844	21.7	15.3	0.5	9.9	17.6	29.6	92.7
Ethnic density (%)	32,844	13.8	18.7	0.0	2.3	5.2	16.7	99.3
Distance (km)	32,844	4.7	4.3	0.04	2.0	3.5	6.0	76.4
Population	32,844	1,666.3	363.6	523	1,446	1,598	1,800	9,551
Pop density (pop/km ²)	32,844	4,423.7	4,506.0	2.5	1,266.8	3,523.7	5,865.3	103,400.0
Non-working-age (%)	32,844	42.6	7.9	1.2	38.9	43.2	47.4	73.6
Participation rate	32,844	1.4	1.5	0.0	0.4	1.0	2.0	15.6

IMD, Index of Multiple Deprivation.

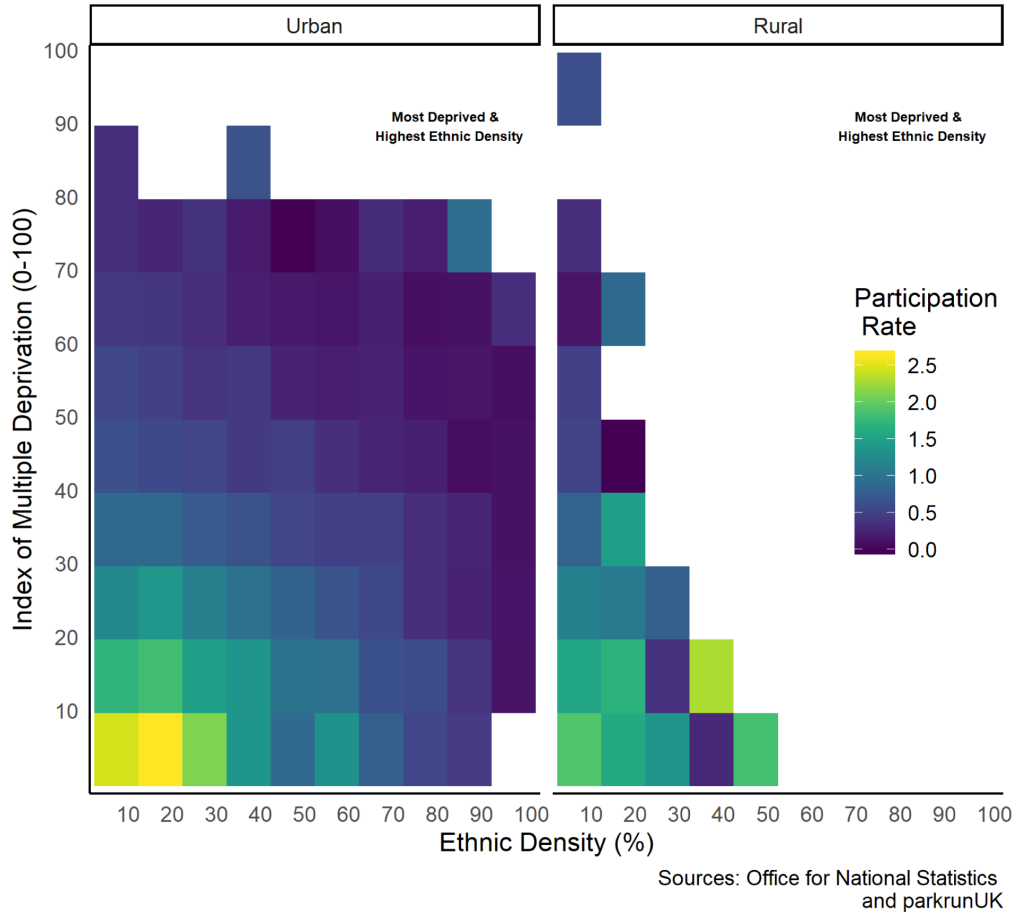


Figure 1. Colour plot for parkrun participation by rural-urban classification, Index of Multiple Deprivation and ethnic density.

Table 3. Poisson log-link generalised linear model results.

	Dependent variable: Finishers		
	Model 1 (IMD)	Model 2 (Ethnic density)	Model 3 (IMD and ethnic density)
IMD score	-0.037* (0.00005)		-0.034* (0.00005)
Ethnic density (%)		-.020* (0.00004)	-0.052* (0.00004)
Pop density (pop/km ²)	-0.107* (0.0004)	-0.118* (0.0004)	-0.070* (0.0004)
Distance (km)	-0.107* (0.0002)	-0.116* (0.0002)	-0.112* (0.0002)
Non-working-age (%)	0.006* (0.00007)	0.002* (0.00007)	-0.001* (0.00007)
Constant	-0.913* (0.005)	-1.068* (0.005)	-0.737* (0.005)
Observations	32,844	32,844	32,844
Log Likelihood	-1,301,151.000	-1,554,894.000	-1,231,308.000
Akaike Inf. Crit.	2,602,312.000	3,109,799.000	2,462,628.000

Note: Std. Error in parenthesis

*p<0.01

IMD, Index of Multiple Deprivation.

persists after controlling for other area characteristics such as deprivation, access to events and population density. While our previous analysis⁸ showed that participation in parkrun is lower in more deprived communities, the present results suggest that a small part of the negative effect on participation previously attributed to deprivation can actually be attributed to ethnic density. parkrun's vision of creating a "healthier and happier planet by continually breaking down barriers to participation and bringing people together from all walks of life whenever they want to come along" (p.5)¹⁴ has potential to improve both population physical activity and community engagement. Identifying the determinants of participation at the community level is a useful first step, but qualitative work to understand why and how these determinants influence participation is an obvious next step. Replicating this study in several years will enable parkrun to monitor trends in participation from different groups in society, and therefore the effectiveness of efforts to reach minority communities and those living in deprived areas.

Limitations

This analysis is ecological and therefore it is not possible to make conclusions at an individual level without risking an ecological inference fallacy. We have been careful throughout to make conclusions at the level of the LSOA, rather than the individual. Nevertheless, given that the evidence at the individual level points to lower participation in organised sport by those from ethnic minority backgrounds⁹, we think it is likely that the same effect exists at the individual level.

Our dependent variable is the number of finishers by residents of each LSOA. This is a count variable where each walk or run finished is treated equally (e.g. 10 finishes by one person is equal to 10 people completing one event). We cannot draw inferences on the number of people who took part within each LSOA at some point in the year, but instead focus on the total finisher count. We do not expect that this will affect the core finding of the paper.

We use percent non-White-British as a crude proxy for ethnic density, and do not estimate participation by ethnic groups separately. It is possible that there are significant differences between participation rates of different minority ethnic groups. Future analysis could look into which groups are more or less engaged in order to better understand the underlying causes of participation. Furthermore, we controlled for several variables that we thought would influence participation but it is possible that there are other confounding factors that have not been included.

Conclusions

parkrun is already in the process of increasing the number of events in deprived areas of England to encourage participation

from disadvantaged groups. Our findings show, however, that in addition to deprivation and access, ethnic density is another important determinant of participation. Breaking down barriers to engagement in parkrun has the potential to improve overall population physical activity and therefore improve overall health and reduce health inequalities.

Data availability

Underlying data

Zenodo: RobertASmith/DoPE_Public: Determinants of parkrun Engagement v1.0. <https://doi.org/10.5281/zenodo.3596841>¹⁰

This project contains the following underlying data:

- /output (folder contains the cleaned data file in CSV format)
- /raw_data/England_Isoa_2011_centroids (LSOA centroid data in DBF, PRJ, SHP and SHX formats)
- /raw_data/IoD2019_Population_Denominators.csv (Non-working age data)
- /raw_data/IoD2019_Scores.csv (Index of Multiple Deprivation score data)
- /raw_data/LSOA_Ethnicity.csv (Ethnic density data)
- /raw_data/LSOA_Rural_Urban_Classification_2011.csv (Rural-urban classification data)
- /raw_data/Mid-2017_Population_Density.csv (Population density data)
- /raw_data/parkrun_data (location data and number of parkrun finishers in CSV format)

Data are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Software availability

Source code available from: <https://github.com/bitowaqr/DoPE>

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.3596841>¹⁰

License: MIT

Acknowledgements

We would like to thank Mike Graney (Head of Analysis at parkrun) and Chrissie Wellington (Global Head of Health and Wellbeing at parkrun) for providing LSOA parkrun participation data.

References

1. Haake S: **Parkrun: a new model of physical activity for large populations?** *The Sport and Exercise Scientist*. 2018; (57): 18–19.
[Reference Source](#)
2. Stevinson C, Hickson M: **Exploring the public health potential of a mass community participation event.** *J Public Health(Oxf)*. 2013; 36(2): 268–274.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Stevinson C, Hickson M: **Changes in physical activity, weight and wellbeing outcomes among attendees of a weekly mass participation event: a prospective 12-month study.** *J Public Health(Oxf)*. 2019; 41(4): 807–814.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Reece LJ, Quirk H, Wellington C, *et al.*: **Bright spots, physical activity investments that work: Parkrun; a global initiative striving for healthier and happier communities.** *Br J Sports Med*. 2019; 53(6): 326–327.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Goyder E, Edmonds C, Sabey A, *et al.*: **What factors predict participation in a mass community physical activity programme? the case of the five sheffield parkruns.** *J Epidemiol Comm H*. 2018; 72(1): A75.
[Publisher Full Text](#)
6. Fullagar S, Petris S, Sargent J, *et al.*: **Action research with parkrun uk volunteer organisers to develop inclusive strategies.** *Health Promot Int*. 2019; daz113.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Cleland V, Nash M, Sharman MJ, *et al.*: **Exploring the health-promoting potential of the parkrun phenomenon: What factors are associated with higher levels of participation?** *Am J Health Promot*. 2019; 33(1): 13–23.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Schneider PP, Smith RA, Bullas AM, *et al.*: **Where should new parkrun events be located? modelling the potential impact of 200 new events on socio-economic inequalities in access and participation.** *medRxiv*. 2019; 19004143.
[Publisher Full Text](#)
9. Rowe N, Champion R: **Sport participation and ethnicity in england. Headline findings.** Sport England, London. 2000.
[Reference Source](#)
10. Smith R: **RobertASmith/DoPE_Public: Determinants of parkrun Engagement v1.0.** 2020.
<http://www.doi.org/10.5281/zenodo.3596841>
11. R Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2018.
[Reference Source](#)
12. Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** Springer-Verlag New York, 2016.
[Reference Source](#)
13. Bibby P, Brindley P: **The 2011 rural-urban classification for small area geographies: a user guide and frequently asked questions (v1. 0).** *Government Statistical Service*. 2013.
14. Cutforth C: **So much more than a run in the park.** *Leisure Rev*. 2017; 1–6.
[Reference Source](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 29 June 2020

<https://doi.org/10.21956/wellcomeopenres.17628.r39168>

© 2020 Senn S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stephen Senn 

Consultant Statistician, Edinburgh, UK

I shall approve this revision. Nevertheless, I think that the authors' response to my criticism of the form of Poisson regression that they used is disappointing. The response boils down to agreeing that the technique used is questionable in general but then claiming that it doesn't matter in this particular instance. This claim is very probably true but just possibly false and it would have been an easy matter for the authors to check: far easier than for any reviewer to do so.

I am not, except by occasional necessity, an R programmer (my usual package for analysis is Genstat) however, looking at my copy of Michael Crawley's book it seems that all that is necessary is to use *quasipoisson* as an alternative argument to *poisson* in the *glm* function (See p582)¹. I am baffled as to why the authors did not think it was worthwhile making the effort to do so.

References

1. Crawley M.J: The R book (second edition). *John Wiley*. 2013.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical methodology; medical statistics; drug development; clinical trials; epidemiology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Jun 2020

Robert Smith, University of Sheffield, Regents Court, Sheffield, UK

When responding to reviewer 3's comments on the use of Poisson regression we stated that we did not think that there was any way the alternative method would change the results, but that others could use the open access code and data to test this. We are keen advocates of open

science and so hoped that others would do this (in fact they already have at several R-hackathon events).

The reviewer has since responded that this is disappointing, and that we should have re-run the analysis, We agree - simply making code and data publicly available does not equal full transparency, particularly when people are trained in different programming languages. We therefore present the results of models 1, 2 and 3 (link below) using a Poisson (P) and QuasiPoisson (QP) regression model. As suspected, the changes do not affect the findings of the study.

We once again thank Reviewer 3 for their useful comments, and apologize for not re-running this relatively simple change to the analysis in the first instance.

https://github.com/SchHARR-PHEDS/DoPE_Public/blob/master/report/QuasiPoisson.pdf

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 June 2020

<https://doi.org/10.21956/wellcomeopenres.17628.r39166>

© 2020 Grunseit A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Anne Grunseit 

¹ The Australian Prevention Partnership Centre, NSW, Australia

² Sydney School of Public Health, University of Sydney, Sydney, Australia

No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Social epidemiology and evaluation in the prevention of lifestyle-related chronic disease.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 18 June 2020

<https://doi.org/10.21956/wellcomeopenres.17628.r39167>

© 2020 Ozakinci G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gozde Ozakinci 

School of Medicine, University of St Andrews, St Andrews, UK

The authors have addressed my query.

Competing Interests: One of the co-authors (Helen Quirk) and I are talking about collaborating on some parkrun data analyses that parkrun collected through a PhD student.

Reviewer Expertise: Health psychology; community physical activity participation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 30 April 2020

<https://doi.org/10.21956/wellcomeopenres.17155.r38442>

© 2020 Senn S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stephen Senn 

Consultant Statistician, Edinburgh, UK

The results of this interesting study are nicely presented and generally well discussed. There are three aspects of the statistical analysis that may be criticised.

First, the authors have used Poisson regression, pointing out that this is commonly used for count data. However, the validity of a Poisson model relies on the assumption that bedrock variability has been reached and this in turn requires that a complete and correct model incorporating all relevant factors has been employed. Furthermore, the Poisson model is a single parameter model with variance equal to expectation. This means that, unlike the Normal model, there is no further play in the model to allow for hidden covariates. This is usually dealt with in one of two ways by modellers. The first is to incorporate a hidden 'frailty' or 'proneeness' parameter. If this is assumed to follow a gamma distribution, then, integrating this out leads to a negative binomial model. This is a two parameter model that can thus allow variances to be greater than predicted by expectation. The second is to check the residual deviance and compare this to the degrees of freedom. The ratio of one to the other then gives a factor by which variances of estimates should be inflated to allow for lack of fit due to hidden random factors. I found no discussion of this point in the paper so can only assume that simple Poisson regression was used, in which case it is likely that the quoted standard errors are too small (See Senn¹ p13 for a discussion).

The second point is that population should perhaps have been used as an offset in the model (see McCullagh and Nelder² p206). Opinions might differ as to how appropriate this is but I would have expected to see it discussed.

The third point is that in controlling for measures of deprivation the authors are asking the question 'given equal deprivation is ethnicity predictive of participation?'. This, is a partial "effect". It may underestimate the role of ethnicity since part of this may be via a tendency to suffer greater deprivation. I am not suggesting that the authors' chosen analysis is inappropriate in controlling for these factors; I am just suggesting that it merits discussion.

References

1. Senn S.J.: Modelling in drug development. *Wiley*. 2011. 35-49
2. McCullagh P, Nelder J.A.: Generalized Linear Models (Second ed.). *Chapman and Hall*. 1989.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: I have recently taken an honorary appointment at the University of Sheffield, however do not know or work with the authors of this article, and believe I am able to write an impartial and objective review.

Reviewer Expertise: Statistical methodology; medical statistics; drug development; clinical trials; epidemiology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 05 Jun 2020

Robert Smith, University of Sheffield, Regents Court, Sheffield, UK

Thank-you very much for reviewing this paper. It is fantastic that we have been able to get peer review on this paper in such a short period of time, so that the findings can have immediate impact.

Our responses to your three points are below:

1) On the validity of a poisson regression model in this instance.

We used a simple Poisson regression model and acknowledge that the standard errors around coefficient point estimates might be underestimated. However, with a sample size of more than 32,000 LSOAs, this problem seems theoretical, as the standard errors are very small. In model 3, for example, the point estimate for 'ethnic density' is -0.052 , and the respective standard error is 0.00004 . To us it seems unlikely that using a negative binomial model would have any relevant effect on the parameter uncertainty and/or the interpretation of the results. However, that being said, we would be very pleased if somebody who is interested would like to investigate this further. All data and code used in this paper are here: <https://doi.org/10.5281/zenodo.3596841>.

2) On population offset:

Thank you for pointing this out – we did indeed use population as an offset variable in the Poisson model, but failed to report this properly in the method section. The respective paragraph has been revised in the following way:

We then used Poisson regression models, commonly used when working with count data, to estimate the relationship between ethnic density, deprivation and parkrun participation, controlling for potential confounding variables including: population density, (population,) age and distance to nearest parkrun event. The LSOA's total population was used as an offset variable.

() = removed.

3) On the partial effect issue:

This is an important point: deprivation might be endogenous, i.e. on the LSOA level, there might be a 'flow of causality' from ethnic density to the level of deprivation. In this case, the effect of ethnic density on participation would be underestimated, as the effect would be partly (and falsely) attributed to deprivation. While controlling for this effect in the statistical model would be a challenge with the data we have, we fully agree with the reviewer that this point deserve mentioning, and thank him for his thoughtful suggestion. The following sentence was added to the discussion:

Finally, it can be assumed that there are some causal relationships between the predictors in our model (e.g. between percentage working age or ethnic density and Deprivation). Future studies should consider conducting mediation analysis, to further disentangle their direct and indirect effects.

We now have participation data for every year from 2010-2019. While it is unlikely that we will be able to solve this specific issue (ethnic density and deprivation don't vary much year by year) we hope to better understand the determinants of parkrun using this more detailed data.

Competing Interests: No competing interests were disclosed.

Reviewer Report 09 March 2020

<https://doi.org/10.21956/wellcomeopenres.17155.r37796>

© 2020 Ozakinci G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gozde Ozakinci 

School of Medicine, University of St Andrews, St Andrews, UK

This is a much needed analysis of potential differences in participation in parkrun. parkrun's popularity raises questions about health inequalities and these sophisticated analyses help us examine the different role that ethnic density and deprivation may play.

I can't comment on the statistical analyses as they are beyond my expertise. The focus of the analyses is the total count of finishers. However, it's not clear to me why a more 'nuanced' approach has not been possible - for instance is it possible to analyse the data in a way that shows the impact of 'parkrun tourism'. Is it possible that this inflates the numbers in some way? By the way, I don't think this will change the conclusions but just pointing it out to encourage authors to give more rationale.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: I'm in the process of developing a collaboration with Helen Quirk.

Reviewer Expertise: Health psychology; community physical activity participation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 04 May 2020

Robert Smith, University of Sheffield, Regents Court, Sheffield, UK

Thank-you for the nice comments on the paper. We appreciate you taking the time to review. To answer your question: The number of finishers is aggregated by geographic

area in which the finisher lives, therefore parkrun tourism should not influence the results.

For example: if I live in an affluent area and travel to a deprived area to do a parkrun this counts towards runs done in affluent areas, not in deprived areas. So, I think you are correct in saying that it would not change the conclusions.

However, what would be interesting in this case is to understand why relatively local parkrun tourism occurs (do people not go to their nearest parkrun because another is more pleasant).

***Competing Interests:* NA**

Reviewer Report 09 March 2020

<https://doi.org/10.21956/wellcomeopenres.17155.r37608>

© 2020 Grunseit A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Anne Grunseit 

¹ The Australian Prevention Partnership Centre, NSW, Australia

² Sydney School of Public Health, University of Sydney, Sydney, Australia

Comments

This paper describes an ecological study of the reach of parkrun in terms of ethnic diversity in parkrun. The paper addresses one of the outstanding issues in the evidence base on this physical activity phenomenon and uses a novel (to parkrun research) method to investigate it. The paper is essentially well-written, has interesting and useful findings and has a sound methodological approach. The question of the effect of clustering within the definition of the participation measure is my only substantive concern, and just requires explanation (see below). Otherwise I have a number of minor comments which should be considered before accepting this article for indexing.

Introduction

Par 1, line 1: I don't think "collection" is the right word – it implies that they are in one place when parkrun's main asset is that it is disseminated. A small wording revision should address this.

Par 4, last sentence: word missing: density would have...

It might be helpful to have a line about physical activity rates among ethnically diverse populations here – sport is not the only form of physical activity and therefore you need to demonstrate that total activity, which is what matters for health, is also lower than for other groups/communities.

Methods

Under data sources, could the authors please indicate whether "finishers" were unique or just a total count ignoring repeat participation (this is mentioned only in the limitations but should be earlier). The authors should describe what the potential implications are of this for the analysis and interpretation – they do

mention that they do not expect it change the results in the limitations but do not provide the rationale for such a conclusion. It also begs the question why they did not use unique persons because it would be possible to do this with parkrun data. In other words they should explain why they chose to operationalise participation this way.

Percentage working age – is there any reason why the authors chose this particular variable for age and how does it relate to the objectives of the analysis?

LSOAs - Could the authors describe why this particular level of spatial classification was used – parkruns draw on varying areas depending on population density but also the proximity to other parkruns. Could they also state whether there were any cases where two or more SOAs were equidistant and if so how were they allocated.

Data analysis – the authors should describe what assumptions for poisson regression were tested (over-dispersion for example). They should also describe how age was operationalised.

Results

The authors talk about ethnic diversity but do not give the reader much idea about what ethnicities this covers in these areas. The authors should describe this somewhere (intro, methods, results) to give the non-UK reader some further context.

Par 1: third last line: Remove “the parkrun distance” as it is not relevant as such to the point being made.

Par 2: The result for age has been reversed in the results from how it was described in the methods which actually makes it more difficult to understand. Is there any reason why you talk about % non-working age rather than % working age? At the very least it should be consistent between methods and results.

Par 3: you make reference to major areas and urban minor areas but have not defined this anywhere. Either here or in the methods would be suitable.

Par on Model 3: As I read the table, despite the attenuation of the effects for IMD and Ethnic density they remained significant in the model – should be explicitly stated in the text.

Limitations

The authors should also note that this research was conducted in one country and the associations may be different in other countries with different geo-demographic patterns and parkrun density.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: I am actively engaged in parkrun research and from time to time submit approval requests to the parkrun Global Research Board of which a number of the authors are a member.

Reviewer Expertise: Social epidemiology and evaluation in the prevention of lifestyle-related chronic disease.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 04 May 2020

Robert Smith, University of Sheffield, Regents Court, Sheffield, UK

Thank-you for taking the time to review and for the many useful comments (which I overwhelmingly agree with).

Introduction

Par 1, line 1: I don't think "collection" is the right word – it implies that they are in one place when parkrun's main asset is that it is disseminated. A small wording revision should address this.

RS: Agreed - changed to use network!

Par 4, last sentence: word missing: density would have...

RS: Happy to change this.

It might be helpful to have a line about physical activity rates among ethnically diverse populations here – sport is not the only form of physical activity and therefore you need to demonstrate that total activity, which is what matters for health, is also lower than for other groups/communities.

RS: We state that "non-White-British individuals in England are less likely to be physically active".

Methods

Under data sources, could the authors please indicate whether "finishers" were unique or just a total count ignoring repeat participation (this is mentioned only in the limitations but should be earlier).

RS: Agreed, this is important for clarity and we have updated this. The number is the total number of finishers (the data we are provided by parkrun is aggregated number of runs by area in order to maintain anonymity). It would be really nice to do with individual level data - but then we would not be able to make the data open access.

The authors should describe what the potential implications are of this for the analysis and

interpretation – they do mention that they do not expect it change the results in the limitations but do not provide the rationale for such a conclusion. It also begs the question why they did not use unique persons because it would be possible to do this with parkrun data. In other words they should explain why they chose to operationalise participation this way.

RS: Agreed, we have updated the publication to state why it is not possible to do this with the data we have, and also that this limitation may causes biases in the results.

Percentage working age – is there any reason why the authors chose this particular variable for age and how does it relate to the objectives of the analysis?

RS: This variable is available in the ONS dataset and was used to control for the effect of having care homes in one area. We wanted to try and control for the effect of having a large community of particularly old people (e.g. a care home or nursing facility) in one LSOA.

LSOAs - Could the authors describe why this particular level of spatial classification was used – parkruns draw on varying areas depending on population density but also the proximity to other parkruns. Could they also state whether there were any cases where two or more SOAs were equidistant and if so how were they allocated.

RS: The LSOA is commonly used as it is the most detailed level of statistic available from the ONS. There were no cases where a LSOA centroid was equidistant.

Data analysis – the authors should describe what assumptions for poisson regression were tested (over-dispersion for example). They should also describe how age was operationalised.

RS: We are in the process of responding to reviewer 3 whose main focus was on these issues. Age was operationalised as a simple percentage non-working age.

Results

The authors talk about ethnic diversity but do not give the reader much idea about what ethnicities this covers in these areas. The authors should describe this somewhere (intro, methods, results) to give the non-UK reader some further context.

RS: This is defined as the percentage of the population who are non-white-British. We didn't look into specific ethnicity, although it would be possible using the data from the ONS.

Par 1: third last line: Remove “the parkrun distance” as it is not relevant as such to the point being made.

RS: Agreed, made change.

Par 2: The result for age has been reversed in the results from how it was described in the methods which actually makes it more difficult to understand. Is there any reason why you talk about % non-working age rather than % working age? At the very least it should be consistent between methods and results.

RS: This is actually a typo in data sources (we have now changed this to match).

Par 3: you make reference to major areas and urban minor areas but have not defined this anywhere. Either here or in the methods would be suitable.

RS: We have updated the data sources to include a description of rural-urban classification.

Par on Model 3: As I read the table, despite the attenuation of the effects for IMD and Ethnic density they remained significant in the model – should be explicitly stated in the text.

RS: Agreed, updated.

Limitations

The authors should also note that this research was conducted in one country and the associations may be different in other countries with different geo-demographic patterns and parkrun density.

RS: Agreed, this point have now been added, along with a call to replicate in other countries. We are particularly keen to see this work replicated and so have made all data and code open access - researchers in other countries with access to that country's ONS equivalent data could easily replicate this work.

Competing Interests: NA