---

bibliography: bibliography.bib

editor\_options:

chunk\_output\_type: console

output:

pdf\_document: default

word\_document: default

---

# Does ethnic density influence community participation in local running events?: a case of parkrun

\*Smith R^1^,Schneider PP^1^, Bullas A^2^, Bayley T^1^, Haake SSJ^2^, Brennan A^1^, Goyder E^1^\*

<br>

\begingroup\small

\*^1^School of Health and Related Research, University of Sheffield\*

\*^2^Sheffield Hallam University Advanced Wellbeing Centre\*

\endgroup

<br>

\_\_Keywords\_\_: `parkrun`, `Participation`, `Physical Activity`, `Deprivation`, `Ethnic density`.

\_\_Intended Journal\_\_: Journal of Racial & Ethnic Health Disparities, Sports Medicine Open,Behavioural and Social Sciences, Journal Of Epi & Comm Health.

\*\*\*\*

### Thumbnail Sketch

#### What is already known on this subject?

parkrun organise weekly 5km running and walking events at parks across the world. Recent research has shown that despite equitable geographical access to parkrun events in England, participation is much lower in more deprived areas.

#### What this study adds?

This study uses regression modelling techniques to better understand the relative influence of geographical access, deprivation and ethnic density on parkrun participation rates in local communities. It finds that areas with higher ethnic density tend to have lower participation rates, even when controlling for deprivation.

#### Policy implications

Identifying why particular communities are less likely to engage in parkrun, and finding ways to improve participation from these commmunities is likely to both improve overall population health and reduce inequalities.

\*\*\*\*

```{r setup, include=FALSE, echo=FALSE}

knitr::opts\_chunk$set(echo = FALSE,

comment = NA,

warning = FALSE,

error = FALSE,

message = FALSE,

tidy = TRUE)

# knitr::opts\_knit$set(root.dir = 'C:/Users/Robert/Google Drive/Other Projects/PARKRUN/DoPE/Dissemination')

```

# Abstract

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.

This paper expands on previous findings by investigating ethnic disparities in parkrun participation. We combined geo-spatial data available through the ONS 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.

We find that areas with higher ethnic density have lower participation rates. This effect is independent of deprivation. An opportunity exists for parkrun to engage with these communities and reduce potential barriers to participation.

# 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 [@haake2018parkrun; @stevinson2013exploring], and there is some evidence that it has increased overall physical activity levels in participants [@stevinson2018changes]. Overall, there is a consensus that parkrun has huge public health potential [@reece2019bright].

However, qualitative research from Sheffield [@goyder2018p2] and more broadly the United Kingdom [@fullagar2019action] 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" [@goyder2018p2].

Undertaking quantitative analysis of the determinants of participation in parkrun is therefore long overdue. Aside from a single previous study from Australia [@cleland2019exploring], 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 rates [@schneider2019]. The analysis was able to quantify, for the first time, how participation in parkrun varied in different communities in England. However, the analysis was interested only in the relationship between participation, access and deprivation and did not consider ethnic density as a potential determinants of participation in parkrun. Yet, evidence from survey data suggest that non-White-British individuals in England are less likely to be physcially active, and to engage in sport in general [@rowe2000sport]. We thus hypothesised that on the community level, the proportion of individuals from racial/ethnic minority groups could have an independent, negative effect on the parkrun participation rate.

# Methods

Data was obtained from multiple sources at the Lower layer Super Output Level (LSOA). There are 32,844 LSOAs in England, each of which is a geographical area containing around 1,000-3,000 people.

[parkrunUK](https://www.parkrun.org.uk/) provided data on the number of parkrun participants from each LSOA in England between the 1st January and 10th December 2018. We also used parkrun event location data, which are publicly available on the [parkrun UK website](https://www.parkrun.org.uk/).

The rest of the data, including Index of Multiple Deprivation (IMD) Score, Ethnic Density, Rural-Urban Classification, Population Density, Percent Working Age and LSOA centroids were obtained from the Office of National Statistics (ONS). Full sources are listed in the table below, and all ONS data is provided open source on the author's GitHub page.

| Variable | Description | Source |

| :----- |:--------| :---------|

| run\_count | number of runs from each LSOA in England between 1st January and 10th December 2018 | parkrunUK |

| imd | IMD scores for each LSOA | [ONS](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/835115/IoD2019\_Statistical\_Release.pdf) |

| total\_pop | total number of individuals in each LSOA | [ONS](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates). |

| pop\_density | population density for each LSOA | [ONS](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareapopulationdensity) |

| rural\_urban | Rural-Urban Classification | [ONS](https://www.gov.uk/government/statistics/2011-rural-urban-classification-of-local-authority-and-other-higher-level-geographies-for-statistical-purposes) |

| perc\_bme | Ethnic Density: percent of population non-white-british | [ONS](https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/2011censussmallpopulationtablesforenglandandwales) |

| mn\_dstn | distance from LSOA centroid to nearest parkrun | derived from [ONS](http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx)|

| perc\_non\_working\_age | derived from ONS data on age-groups in each LSOA | [ONS](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates) |

| run\_rate | derived from run\_count and LSAO populations | derived |

After merging these datasets we had detailed data on 32,844 LSOAs, including participation and several characteristics of the LSOAs which we hypothesised may influence participation. Since previous work has found corelations between participation and deprivation, distance to nearest event, and population density we included all of these variables. We also extended the analysis to include ethnic density (we use the percent of the population that reported being non-White-British as a proxy for ethnic density) and the percent of the population of working age. We are interested in ethnic density as we hypothesised that areas with higher ethnic density would have lower participation rates, all else being equal. We included the percent of the population that is working age as a control to limit for the effect of populations heavily skewed toward older people (e.g. care homes), or very young people (e.g orphanages/immediately around special needs schools)). Since participation in parkrunUK is dominated by those aged 20-60 [@haake2018parkrun] we felt this was justified.

We first studied the bivariate Pearson correlations between the dependent variable (parkrun participants per week per 1,000 population) and all indepdent variables (x,y,z). Results are visually illustrated, using the correlation plots [@corrplot2017] and (stratified) heat maps. We then fitted a multivariable regression model to investigate the independent effects of the predictors on parkrun participation. A Poisson distribution with a log-link and with the LSOA total population as an offset variable was used to model parkrun participation as a rate (runners per LSOA population). Model fit was assessed using Pseudo R^2, based on quasi-likelihood functions (Zhang, 2017).<!--rsq package: Zhang, D. (2017) rsq: An R package to calculate the R-squared for generalized linear models. In preparation.--> All analyses were conducted in R [@base\_r].

# Results

## Descriptive Statistics

Participation in parkrun varies across LSOAs. Around half of all communities (LSOA) average less than 1 finisher per week per 1000 people. Approximately a quarter average between 1 and 2 runs, and around an eighth between 2 and 3 runs. There is considerable variation in ethnic density, with most LSOA having a large majority of White-British residents, and few areas having over 50% non-White-British residents. Deprivation is positively skewed, meaning that most areas are not deprived, with a few very deprived areas. Finally, around 70% of LSOAs are within 5km, a parkrun, of a parkrun. Again this is positively skewed with most LSAO being within 3-4km.

```{r, echo = F,fig.height=5, fig.width=8}

library(reshape2)

library(dplyr)

library(tidyverse)

library(ggplot2)

library("viridis")

#install.packages("corrplot")

library(corrplot)

library(ggplot2)

library(fitdistrplus)

#setwd('C:/Users/Robert/Google Drive/Other Projects/PARKRUN/DoPE/Dissemination')

#df <- read.csv("./output/lsoa\_df.csv") %>%

# mutate(run\_rate = run\_count/total\_pop \* 1000/52)

# distribution of

df <- read.csv("../output/lsoa\_df.csv") %>%

dplyr::select(run\_count,imd, perc\_bme, mn\_dstn,

total\_pop, pop\_density,perc\_non\_working\_age) %>%

mutate(run\_rate = run\_count/total\_pop \* 1000/52)

# distribution of run rates

plot\_part <- (ggplot(data=df, aes(run\_rate)) +

geom\_histogram(aes(y =..density..),

breaks=seq(0, 800/52, by = 1),

col="darkgreen",

fill="green",

alpha = .2) +

#geom\_density(col=2) +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5))+

labs(title="Participation") +

labs(x="Finishers per 1000 persons", y = "Density"))

# distribution of imd rates

plot\_imd <- (ggplot(data=df, aes(imd)) +

geom\_histogram(aes(y =..density..),

col="darkgreen",

fill="green",

alpha = .2) +

#geom\_density(col=2) +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5))+

labs(title="IMD Score") +

labs(x="IMD Score", y = "Density"))

# disetribution of Ethnic Density

plot\_ed <- (ggplot(data=df, aes(perc\_bme)) +

geom\_histogram(aes(y =..density..),

col="darkgreen",

fill="green",

alpha = .2) +

#geom\_density(col=2) +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5))+

labs(title="Ethnic Density") +

labs(x="% non-WhiteBritish", y = "Density"))

# distribution of distance to nearest event

plot\_dist <- (ggplot(data=df, aes(mn\_dstn)) +

geom\_histogram(aes(y =..density..),

col="darkgreen",

fill="green",

alpha = .2) +

#geom\_density(col=2) +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5))+

xlim(c(0,40))+

labs(title="Nearest event") +

labs(x="Distance (km)", y = "Density"))

library(gridExtra)

grid.arrange(plot\_part,plot\_imd,plot\_dist,plot\_ed,ncol=2,nrow=2)

```

\newpage

## Correlation Matrix

There is a negative correlation between particpation (run\_count) and: deprivation (imd), distance to nearest parkrun (mn\_distance), population density (pop\_density) and ethnic density (perc\_bme). Ethnic density is strongly postively correlated with population density, negatively correlated with percent non-working age, and moderately positvely correlated with IMD suggesting that areas with higher ethnic density are more densely populated overall, more deprived and have fewer older people.

```{r, echo = F,fig.height=5, fig.width=8}

# setwd('C:/Users/Robert/Google Drive/Other Projects/PARKRUN/DoPE')

## CORRELATION MATRIX IMD ONLY

cor\_mat <- read.csv("../output/lsoa\_df.csv") %>%

dplyr::select(run\_count,imd, perc\_bme, mn\_dstn,

total\_pop, pop\_density,perc\_non\_working\_age) %>%

cor

rownames(cor\_mat) <- colnames(cor\_mat) <- c("Participation","IMD","Ethnic Density","Access","Total Pop","Pop Density","% Non Working") #substr(colnames(cor\_mat),1,20)

corrplot(corr = cor\_mat,

addCoef.col = "black",

type = "upper")

title = "Correlation of Variables")

```

\newpage

```{r, eval = F,echo = F,fig.height=5, fig.width=8}

## Association plot

df <- read.csv("../output/lsoa\_df.csv") %>%

dplyr::select(run\_count,imd, perc\_bme, mn\_dstn,

total\_pop, pop\_density,perc\_non\_working\_age)

## hist bme

# ggplot(df) +

# geom\_histogram(aes(perc\_bme),fill="cyan",alpha=0.5,col="blue")

# boxplots run\_count~ bme\_10

# bme\_10 = cut(df$perc\_bme,breaks = quantile(df$perc\_bme,probs=seq(0,1,by=0.1)),include.lowest = T)

# ggplot(df) +

# geom\_boxplot(aes(y=run\_count,fill= bme\_10,x=bme\_10)) +

# ylim(c(0,500)) +

# theme()

## Association plot

# ggplot(df) +

# geom\_point(aes(x=perc\_bme\*100,y=run\_count),col="purple",alpha=0.3,size=0.5) +

# geom\_smooth(aes(x=perc\_bme\*100,y=run\_count)) +

# labs(y="Average weekly runs per 1000 people", x = "% ethnic minorities") +

# ylim(c(0,750)) + # outliers excluded

# theme\_minimal()

#

#

# ggplot(df) +

# geom\_point(aes(x=perc\_bme\*100,y=imd),col="darkorange",alpha=0.3,size=0.5) +

# geom\_smooth(aes(x=perc\_bme\*100,y=imd)) +

# labs(y="IMD score", x = "% ethnic minorities") +

# theme\_minimal()

#

#

# ggplot(df) +

# geom\_point(aes(x=imd,y=run\_count),col="cyan",alpha=0.4,size=0.5) +

# geom\_smooth(aes(x=perc\_bme\*100,y=run\_count)) +

# labs(y="Average weekly runs per 1000 people", x = "IMD") +

# ylim(c(0,750)) + # outliers excluded

# theme\_minimal()

# joint smooth association plot

bme\_5 = cut(df$perc\_bme,breaks = quantile(df$perc\_bme,probs=seq(0,1,by=0.8)),include.lowest = T)

ggplot(df) +

geom\_point(aes(x=imd,y=run\_count),col="darkgray",alpha=.3,size=.5) +

geom\_smooth(aes(x=imd,y=run\_count,col=bme\_5)) +

xlab("Multiple deprivation") +

ylim(c(0,500)) +

ylab("Average weekly runs per 1000 people") +

theme\_minimal() +

scale\_color\_manual(values= c(2,3,4,5,6),name="% bme quintiles",labels=c("Lowest","low","median","high","highest"))

```

\newpage

## Colour plot for rural/urban.

The colour plots below show the participation levels for LSOA by deprivation and ethnic density for Urban Major (XY pop/km^2), Urban Minor (XY/km^2) and Rural areas (XY/km^2). Yellow indicates high, green indicates moderate, and dark blue indicates low levels of participation.

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 and there are therefore many confounding factors.

```{r, echo=F,fig.height=5, fig.width=8}

# RUN RATE

df <- read.csv("../output/lsoa\_df.csv")

lsoa\_ruralurban <- read.csv("../raw\_data/LSOA\_Rural\_Urban\_Classification\_2011.csv",stringsAsFactors = F) %>%

mutate(urban = replace(RUC11CD,RUC11CD %in% c("A1","B1"),"Urban Major"),

urban = replace(urban, urban %in% c("C1","C2"),"Urban Minor"),

urban = replace(urban, urban %in% c("D1","D2","E1","E2"),"Rural"),

urban = factor(urban,ordered = TRUE,levels = c("Urban Major","Urban Minor","Rural"))) %>%

dplyr::select(code,urban) #%>%

# mutate(#urban = replace(urban,urban==TRUE,"Urban"),

# urban =replace(urban,urban==FALSE,"Rural"))

df <- merge(df,lsoa\_ruralurban) %>%

mutate(run\_rate = run\_count/total\_pop,

imd\_dec = cut(x = imd,

breaks = seq(0,100,10), # quantile(imd,seq(0,1,0.1)),

ordered\_result = T,

labels = F)\*10,

bme\_dec= cut(x = perc\_bme,

breaks = seq(0,1,0.1), # quantile(perc\_bme,seq(0,1,0.1)),

ordered\_result = T,

labels = F)\*10)%>%

melt(id.vars = c("code","imd\_dec","bme\_dec","urban"),

measure.vars ="run\_rate",

value.name = "run\_rate") %>%

dplyr::select(imd\_dec,bme\_dec,run\_rate,urban)

#df$pop\_density\_bins = cut(df$pop\_density,

# quantile(df$pop\_density,

# probs = c(0,0.25,0.5,0.75,1)),

# include.lowest = T,

# labels = c("Low density","High density")) # c("Lowest density","Low density","High density","Highest density")

df <- aggregate(run\_rate ~ bme\_dec + imd\_dec + urban, #+ #pop\_density\_bins,

data = df,

FUN= "mean")

ggplot(data = df,

aes(as.factor(bme\_dec), as.factor(imd\_dec), fill= run\_rate)) +

geom\_tile()+

scale\_fill\_viridis(discrete=FALSE) +

xlab("Ethnic Density (%)")+

ylab("Index of Multiple Deprivation (0-100)") +

facet\_wrap(~urban, nrow = 2) +

labs(caption="Source: Office for National Statistics")

#NULL

```

\newpage

```{r, eval = F, echo=F,fig.height=5, fig.width=8}

# Colour plot split by distance to nearest event.

# RUN RATE

df <- read.csv("../output/lsoa\_df.csv")

lsoa\_ruralurban <- read.csv("../raw\_data/LSOA\_Rural\_Urban\_Classification\_2011.csv",stringsAsFactors = F) %>%

mutate(urban = replace(RUC11CD,RUC11CD %in% c("A1","B1"),"Urban Major"),

urban = replace(urban, urban %in% c("C1","C2"),"Urban Minor"),

urban = replace(urban, urban %in% c("D1","D2","E1","E2"),"Rural"),

urban = factor(urban,ordered = TRUE,levels = c("Urban Major","Urban Minor","Rural"))) %>%

dplyr::select(code,urban) #%>%

# mutate(#urban = replace(urban,urban==TRUE,"Urban"),

# urban =replace(urban,urban==FALSE,"Rural"))

df <- merge(df,lsoa\_ruralurban) %>%

mutate(run\_rate = run\_count/total\_pop,

imd\_dec = cut(x = imd,

breaks = seq(0,100,10), # quantile(imd,seq(0,1,0.1)),

ordered\_result = T,

labels = F)\*10,

bme\_dec= cut(x = perc\_bme,

breaks = seq(0,1,0.1), # quantile(perc\_bme,seq(0,1,0.1)),

ordered\_result = T,

labels = F)\*10,

mn\_dstn= cut(x = mn\_dstn,

breaks = c(0,5,10000),

labels = c("Under 5K","Over 5K")))%>%

melt(id.vars = c("code","imd\_dec","bme\_dec","urban","mn\_dstn"),

measure.vars ="run\_rate",

value.name = "run\_rate") %>%

dplyr::select(imd\_dec,bme\_dec,run\_rate,urban,mn\_dstn)

#df$pop\_density\_bins = cut(df$pop\_density,

# quantile(df$pop\_density,

# probs = c(0,0.25,0.5,0.75,1)),

# include.lowest = T,

# labels = c("Low density","High density")) # c("Lowest density","Low density","High density","Highest density")

df <- aggregate(run\_rate ~ bme\_dec + imd\_dec + mn\_dstn, # + urban, #+ #pop\_density\_bins,

data = df,

FUN= "mean")

ggplot(data = df,

aes(as.factor(bme\_dec), as.factor(imd\_dec), fill= run\_rate)) +

geom\_tile()+

scale\_fill\_viridis(discrete=FALSE) +

xlab("Ethnic Density (%)")+

ylab("Index of Multiple Deprivation (0-100)") +

facet\_wrap(~mn\_dstn, nrow = 2) +

labs(caption="Source: Office for National Statistics")

#NULL

```

\newpage

## Poisson Model

The results of the poisson regression show that areas with a higher ethnic density have lower participation rates, even when controlling for the effect of deprivation and distance to events. The effect is smaller than deprivation and distance, but still material and significant.

<!-- NOTE: we could show bivariate regression for each of the variables alone? -->

```{r, echo = FALSE,results='asis'}

#install.packages("stargazer")

library(stargazer)

rm(list = ls())

df <- read.csv("../output/lsoa\_df.csv")

df$pop\_density = log(df$pop\_density ) # tranforming pop\_density to log scale

#===

# Model 1: Poisson model imd all as one.

#===

model1 = glm(run\_count ~ imd + pop\_density + mn\_dstn

+ perc\_non\_working\_age + perc\_bme,

data = df,

family = poisson(link="log"),

offset = log(total\_pop))

# stargazer(model1,ci=TRUE, ci.level=0.95)

x = summary(model1)

r2.1 = 1-((x$deviance-length(coef(x)[,1]))/x$null.deviance)

# r2.1

#===

# MODEL 2 - SCALED DATA - allows easier interpretation of beta coefs

#===

# scale the dataframe

scaled\_df <- data.frame(code = df$code,

run\_count = df$run\_count,

scale(df[,-c(1,2,12)]),

total\_pop =df$total\_pop,

stringsAsFactors = F)

# model 2

model2 <- glm(run\_count ~ imd + mn\_dstn +

perc\_non\_working\_age + pop\_density + perc\_bme,

data = scaled\_df,

family = poisson(link="log"),

offset = log(total\_pop))

# stargazer(model2,ci=TRUE, ci.level=0.95)

x <- summary(model2)

r2.2 = 1-((x$deviance-length(coef(x)[,1]))/x$null.deviance) # same fit

# r2.2

#===

# MODEL 3

#===

# model3 <- glm(run\_count ~ imd \* perc\_bme + mn\_dstn,

# data = scaled\_df,

# family = poisson(link="log"),

# offset = log(total\_pop))

#summary(model3)

# x = summary(model3)

# r2.3 = 1-((x$deviance-length(coef(x)[,1]))/x$null.deviance)

#===

# MODEL 4 - JUST IMD IMD, DISTANCE AND %BME

#===

model4 <- glm(run\_count ~ imd + mn\_dstn + perc\_bme,

data = scaled\_df,

family = poisson(link="log"),

offset = log(total\_pop))

# summary(model4)

x = summary(model4)

r2.4 = 1-((x$deviance-length(coef(x)[,1]))/x$null.deviance)

##### summary

stargazer(

model1,model4,model2,

column.labels = c("Original scale","Scaled - min model","Scaled - full model"),

ci=TRUE, ci.level=0.95,

title="Regression Results",header = FALSE)

#### bme percentile model

# bme\_10 = cut(df$perc\_bme,breaks = quantile(df$perc\_bme,probs=seq(0,1,by=0.1)),include.lowest = T)

# model5 <- glm(run\_count ~ imd + mn\_dstn +

# perc\_non\_working\_age + pop\_density + bme\_10,

# data = scaled\_df,

# family = poisson(link="log"),

# offset = log(total\_pop))

#

# summary(model5)

#

# x = summary(model5)

# r2.5 = 1-((x$deviance-length(coef(x)[,1]))/x$null.deviance)

```

```{r}

# Let's discuss whether Venn diagrams are useful here

# they show different and overlapping R2 proportions

# which could be interesting, if we include a bit on

# intersectionality/interaction, maybe?

```

\newpage

# Discussion

Our findings show that areas with higher % bme have lower participation rates. The effect persists after controlling for other area characteristics. <!--a bit more elaboration needed--> While our previous analysis [Scheider et al., 2019](https://www.medrxiv.org/content/early/2019/08/29/19004143.full.pdf) has shown that participation in parkrun is lower in more deprived communities, the present results suggest that some of the negative effect on participation previously attributed to deprivation can actually be attributed to differences in participation by ethnic minorities.

parkrun's mission states that they aim to increase levels of physical activity in deprived communities. Our findings indicate that participation in deprived communities with ethnic density is particularly low. Further research could be undertaken to ascertain trends in participation from different groups in society, allowing parkrun to monitor the effectiveness of their efforts to reach minority communities. More research is needed to understand the barriers to attending parkrun for members of those communities.

# Limitations

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

Our dependent variable is the number of runs by residents of each LSOA. This is a count variable where each walk/run finish is treated equally (e.g. 10 runs by one person is equal to 1 run by 10 people). 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 controlled for several variables which we thought would influence participation, it is possible that there are other confounding factors which have not been included.

# Conclusion

parkrun is already in the process of increasing the number of events in deprived areas in England to encourage more participation from disadvantaged groups. Our findings show, however, that in addition to deprivation and access, ethnic density is another important determinant of participation, which should be taken into account when planning future parkrun event locations. Breaking down barriers to engagement in parkrun has the potential to improve overall population physical activty and therefore improve overall health and reduce health inequalities.

# References