The University Of Sheffield.

This is a repository copy of *Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/142753/

Version: Accepted Version

Article:

Brindley, P.G. orcid.org/0000-0001-9989-9789, Cameron, R. orcid.org/0000-0002-7786-0581, Ersoy, E. et al. (2 more authors) (2019) Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health. Urban Forestry & Urban Greening. ISSN 1618-8667

https://doi.org/10.1016/j.ufug.2019.01.015

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Accepted Manuscript

Title: Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health

Authors: Paul Brindley, Ross W. Cameron, Ebru Ersoy, Anna Jorgensen, Ravi Maheswaran

Please cite this article as: Brindley P, Cameron RW, Ersoy E, Jorgensen A, Maheswaran R, Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health, *Urban Forestry and amp; Urban Greening* (2019), https://doi.org/10.1016/j.ufug.2019.01.015

Full Title: **Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health**

Paul Brindley [a] [*], Ross W. Cameron [a], Ebru Ersoy [b], Anna Jorgensen [a], Ravi Maheswaran [c]

[a] Department of Landscape Architecture, The University of Sheffield, Sheffield, United Kingdom. E-Mails: p.brindley@sheffield.ac.uk, r.w.cameron@sheffield.ac.uk, a.jorgensen@sheffield.ac.uk

[b] Department of Landscape Architecture, Adnan Menderes University, Aydın, Turkey. E-Mail: ebruersy@gmail.com

[c] Public Health GIS Unit, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom. E-Mail: r.maheswaran@sheffield.ac.uk

[*] Author to whom correspondence should be addressed; Department of Landscape Architecture, University of Sheffield, The Arts Tower, Western Bank, Sheffield, S10 2TN, UK; E-Mail: p.brindley@sheffield.ac.uk; Tel.: +44 114 2220615; Fax: +44 114 2220627

Highlights

- We examined associations between greenspace quality and self-reported general health
- We also explored the potential for social media as proxy measures for greenspace quality
- Poor correlation was found between social media data and field survey greenspace quality
- We found no association between social media measures and health outcomes
- Higher levels of cleanliness recorded by the field survey were associated with better general health

1. Introduction

Despite the growing recognition of the health benefits of greenspace, the literature highlights a lack of evidence investigating the role of greenspace 'quality' rather than quantity (or provision) (Akpinar, 2016; De Vries et al., 2003; van den Berg et al., 2007; Wyles et al., 2017). This is most likely due to the difficultly in defining the subjective term of greenspace 'quality' and the time and cost of measuring quality across all greenspaces within a study area in a comprehensive and systematic approach. The digital explosion of data offered by social media is able to provide insights into human behaviour when opinions are collectively aggregated together en masse (Tsakalidis et al., 2016). The aims of this work are two-fold. The first aim is to investigate the possible role of data extracted from

1

social media as a proxy for greenspace quality as measured using field surveys. The second aim is to undertake a population level study to investigate the relationship between health and greenspace quality, using the city of Sheffield, UK, as a case study.

This work forms part of a wider research project (IWUN - http://iwun.uk) which seeks to find out more about how Sheffield's natural environment can improve the health and wellbeing of the city's residents. Sheffield as a city is spatially divided both in terms of deprivation and health – with the geographical divide running north-west to south-east. The 'Fairness On The 83' project reported that average life expectancy fell by 7.5 years for men and 10 years for women along the 83 bus route that runs across the divide (Sheffield Fairness Commission, 2013). Our work in IWUN demonstrates that the more deprived east of Sheffield does not have substantially less greenspace, justifying our enquiry as to whether it's the quality of the greenspace that may have an effect on health outcomes.

There is growing body of research suggesting that exposure to greenspace enhances health and wellbeing (Bosch and Sang, 2017). The reported health benefits of greenspace are broad, and range from the capacity to reduce obesity, to improving educational performance (Lovell, 2016). Other health benefits include promoting mental health (for example by reducing the risk of stress, tendency to psychiatric morbidity, psychological distress, depressive symptoms, clinical anxiety, depression and mood disorders); affecting birth outcomes; influencing physiological health (for example cancer, diabetes, cardiovascular outcomes); improving general health; and ultimately impacting on mortality (Dadvand et al., 2016; Dzhambov et al., 2014; Gascon et al., 2016, 2015; James et al., 2015; Keijzer et al., 2016; Lanki et al., 2017; Lee and Maheswaran, 2011; Lovell, 2016).

Markevych et al., (2017) organised the possible mechanisms for the health benefits of greenspace into three domains: (1) Reducing harm (mitigation) - for example, reducing exposures to environmental stresses: air pollution, noise and heat; (2) Restoring capacities (restoration) - for example, attention restoration and psychophysiological stress recovery; and (3) Building capacities (instoration) - for example, encouraging physical activity and facilitating social cohesion.

Whilst much of the literature, to date, has focused on the connection between the presence / absence (or in other words the amount of greenspace) and health benefits, there is a recognition that the 'quality' rather than the quantity of greenspaces may play an important role in the relationship between greenspace and health (Akpinar, 2016; Akpinar et al., 2016; De Vries et al., 2003; Maas et al., 2006; Ord et al., 2013; Richardson et al., 2010; Richardson and Mitchell, 2010; van den Berg et al., 2007; Vujcic et al., 2018; Wyles et al., 2017). Despite this, Lachowycz and Jones (2013) acknowledge the problem that information about the quality of greenspace is seldom available.

Definitions of greenspace quality differ but broadly deal with levels of maintenance and/or the intrinsic qualities of a space, i.e. the various elements that make up a space, and the activities they support. Definitions of greenspace quality are discussed further in Section 1.2. In this paper we focus on the maintenance dimension of greenspace quality.

The existing literature exploring the role of greenspace quality for improving health has generally focused on social amenity and environmental incivility, such as litter (Wheeler et al., 2015) and has largely centred on the role of physical activity (Akpinar, 2016; Giles-Corti et al., 2005; Hillsdon et al., 2006). The pristineness and aesthetic value of the environment (for example the absence of litter) has also been shown to have beneficial restorative effects (Pretty et al., 2005; Wyles et al., 2016).

The potential mechanisms by which greenspace *quality* may provide health benefits are similar to those associated with greenspace more generally – as previously discussed. The quality of

2

greenspace may affect the number of users of the greenspace – as people will choose whether or not to use the space not only for its features but also the condition of those features (Lee and Maheswaran, 2011). Places in disrepair are less likely to be visited and contribute to a perceived sense of lack of safety (Bedimo-Rung et al., 2005; Maruthaveeran et al., 2014). Conversely, quality greenspaces attract greater numbers of people who are likely to undertake physical activity within the greenspace as well as enjoying opportunities for social contact (Barker et al., 2017). In their socio-ecological framework for the relationship between green space access and health Lachowycz and Jones (2013) have suggested that greenspace quality (defined as maintenance, design, attractiveness) may moderate motivation to use greenspace. Furthermore, quality may affect the visual cues provided by the greenspace, enhancing psychological pathways to restoration and supporting place identity and attachment.

1.1 Previous literature related to the health benefits from greenspace quality

Using individuals from a survey of 4,515 people in England, Wyles et al. (2017) found that respondents recalled greater connectedness to nature and restoration when visiting sites of higher environmental quality. In a New York, USA survey of 44,282 participants, lower park cleanliness (presence of litter, glass, weeds, and graffiti) was associated with higher neighbourhood levels of body mass index (Stark et al., 2014). Despite the potential importance of greenspace quality in the relationship between greenspace and health, Van Den Berg et al. (2015) found that there were "insufficient studies on the quality of green spaces to conduct an evidence synthesis" (p.806). The lack of existing population level (ecological) studies considering the role of greenspace quality when exploring the health benefits of greenspace is widely recognised (Bowler et al., 2010; Hartig et al., 2014; Van Den Berg et al., 2015; Wheeler et al., 2015).

The majority of studies concerning the health benefits associated with greenspace quality have explored the relationship with perceived general health. Three Dutch studies found mixed associations between greenspace quality and perceived general health (Agyemang et al., 2007; Putrik et al., 2015; van Dillen et al., 2012). No association was found for general health by Putrik et al. (2015), who aggregated perceptions of greenspace quality and quantity by asking residents to collectively rate their "quality and availability of green space" (p.51). Mixed results were reported by Agyemang et al. (2007), who explored the health benefits associated with resident dissatisfaction with greenspace quality in their neighbourhood. In a tertile categorisation of greenspace dissatisfaction, the lowest dissatisfaction was associated with the lowest levels of poor health. However, the highest levels of dissatisfaction were found for the centre tertile for poor health (not the highest category as one might expect). van Dillen et al. (2012) found that residents in neighbourhoods with a higher quality greenspace experienced higher levels of perceived general health and less acute health-related complaints in the last 14 days. Both greenspace quality and quantity measures were found to be associated with general health – with the quality measures improving the relationship beyond simply including greenspace quantity on its own.

In a UK study, Wheeler et al. (2015) used a measure of surface water quality as a general indicator of landscape quality – thus demonstrating the general lack of available data relating to greenspace quality. They found, in contrast to expectation, that poorer water quality was associated with better perceived general health in the England and Wales. Greenspace quality was found not to be an adequate predictor of well-being by Zhang et al. (2017), although they did find that residents' perceived quality of greenspaces were significantly associated with neighbourhood satisfaction.

In a Dutch study exploring the relationship between quality and self-reported mental health, De Vries et al. (2013) showed that streetscape greenery quality was significantly associated with better mental health (over and above quantity). In contrast, however, the study by van Dillen et al. (2012) (discussed previously in relation to general health), found that quality of greenspace was not associated with levels of mental health. Whilst a number of studies have explored the associations between quantity of greenspace and mortality, none have investigated its relation with quality of greenspace (Van Den Berg et al., 2015).

These mixed findings from the literature demonstrate that the evidence regarding the role of quality within benefits of greenspace is still inconclusive. Therefore, the first aim of this paper is to add to the literature by assessing the role of greenspace quality in a population level study of self-rated general health in Sheffield, UK.

1.2 Reasons behind the lack of studies exploring the health benefits of greenspace quality

There are a number of factors which might influence the general lack of population level studies investigating the potential health benefits of greenspace quality. Firstly, as previously indicated, the notion of 'quality' can be interpreted in many differing ways. Custodians and managers of greenspace may hold disparate perspectives of quality compared to public users. There is no single definition, merely an acceptance that 'quality' can be defined in numerous ways. To some, it might be an expression of the amount of litter, but to others it might reflect the levels of biodiversity, or relate to the incidence of crime. The remit is broad and diverse.

De Vries et al. (2013) measured quality by scoring levels of accessibility, maintenance, variation, naturalness, colourfulness, clear arrangement, shelter, absence of litter, safety and overall general impression. In contrast, Akpinar (2016) identified the following aspects of quality in urban greenspace: aesthetics, cleanliness, maintenance, largeness, shading, lighting, and openness/visibility. Alternatively, Zhang et al. (2017) quantified quality on a six item scale including measures related to: facilities, amenities, natural features, incivilities, accessibility and maintenance. It is also possible to use the designation of protected environments (official status recognising the site is of particular scientific, ecological, or aesthetic value) to reflect the quality of the site as a measure of greenspace quality (as used by Wyles et al., 2017). Whilst such an approach may be suitable for large scale, national studies, it is unlikely, however, to be appropriate for city-wide population level studies due to a lack of protected environments within urban contexts. For example, of the 902 greenspaces within the urban area of Sheffield (as defined by the Built-Up Area, Office for National Statistics) only 38 consist of protected or designated areas. Their designation tells us nothing about the remaining 864 non-designated greenspaces, which might themselves be high or low 'quality'. As previously mentioned, in this paper we focus on the maintenance dimension of greenspace quality.

A second possible reason for the lack of population level studies in this area relates to the issue that current measures of greenspace quality are derived predominantly from primary audits (also called field surveys) of local environments. These are both time consuming and costly and it is therefore difficult to audit systematically all greenspaces within any substantial geographic extent (e.g. throughout a city). The second aim of this paper is to explore whether potential proxies of greenspace quality might be drawn from alternative sources – in this case through the use of social media data (as discussed in Section 1.3 below).

1.3 Social media as a potential proxy for greenspace quality

Social media is a transformative digital technology that has reshaped public communication. Social media data are increasingly perceived as alternative sources to public surveys (Department for Work and Pensions, 2014; Tsakalidis et al., 2016) and have been used to aid understanding in a range of contexts from rioting (Procter et al., 2013) to museum exhibits (Gerrard et al., 2017).

A number of studies have exploited social media to explore greenspace usage. For example, visitor numbers to American parks have been shown to be reliably estimated by both Flickr photos (Donahue et al., 2018; Sessions et al., 2016) and Tweets from Twitter (Donahue et al., 2018). Internet hosted photography sites (including Flickr, Instagram and Panaramio) have also been used to capture the aesthetic landscape value and perceptions of users (Dunkel, 2015; Figueroa-Alfaro and Tang, 2017; Zanten et al., 2016).

The potential merit of applying such an approach in Sheffield was highlighted by preliminary exploration, as shown in Fig. 1, which demonstrates the different words used within the titles and captions of Flickr photographs to describe two contrasting Sheffield greenspaces. Many of the words associated with the first greenspace (Fig. 1a) include references to nature, in stark contrast to the other greenspace, which was more likely to be linked with words related with crime (Fig. 1b). It is within this context that our paper explores the potential for certain social media data (extracted from Flickr and Twitter) to proxy measures of greenspace quality.

Measures extracted from social media include (1) the number of photographs taken within greenspaces and uploaded to Flickr; (2) the sentiment of those photograph captions; and (3) the number of geo-referenced Tweets from Twitter within greenspaces. Quality in this context might be considered as an expression of the number of greenspace users (expressed by the number of social media users in the area), feeling comfortable enough to use social media within the area, and the assumption that users were presumably inspired to take a photograph of something of interest.

1.4 Aims of the study

The aims of this work are two-fold. The first aim is to explore the potential for data extracted from social media to proxy greenspace quality measured using field surveys. The second aim is to undertake a population level study to investigate the relationship between poor general health and greenspace quality.

2. Methods

2.1 Case study

This study was conducted within Sheffield, UK and forms part of a wider case study exploring how Sheffield's natural environment can improve the health and wellbeing of residents (http://iwun.uk). According to the 2011 Census, the population of Sheffield Local Authority is 552,698. Our work utilised general health data from the UK 2011 Census of Population, greenspace quality data from Sheffield City Council, potential proxy measures of greenspace quality from social media (Flickr and Twitter) and a wide range of other data to account for confounders. Data were collected or aggregated to Lower-layer Super Output Areas (LSOAs), which are a geographic unit commonly used for reporting small area statistics in England and contain an average population of approximately 1,500. There are 345 LSOAs in Sheffield.

5

2.2 Health data

Self-reported general health was obtained from the 2011 UK Census at the LSOA scale. General health is a self-assessment of a person's general state of health. People were asked the question "How is your health in general?" and could respond with the following answers: very good, good, fair, bad or very bad. The assessment was not based on a person's health over any specified time period. The poor health measure was constructed by aggregating the 'bad' and 'very bad' health categories. According to the 2011 Census, 6.2% of Sheffield's population reported poor health. Indirect standardisation was undertaken for broad age (0 to 15; 16 to 24; 25 to 34; 35 to 49; 50 to 64; 65 to 74; 75 to 84; and 85 and over) and sex categories. Self-reported general health from the UK Census has been used within a number of similar population level studies (Brindley et al., 2018; Mitchell and Popham, 2007; Wheeler et al., 2015), is considered a good predictor of morbidity and mortality (Desalvo et al., 2006; Idler and Benyamini, 1997) and has been shown to be a reliable measure of objectively measured health outcomes (Kyffin et al., 2004; Mavaddat et al., 2011; Short et al., 2009).

2.3 Greenspace data

Sheffield City Council supplied spatial data relating to 945 greenspaces within the study area, 850 of which were included in a quality assessment undertaken in 2008 as part of compliance with former Governmental guidance – Planning Policy Guidance note 17 (PPG17: 'Open Space, Sport and Recreation'). Manual checking of the 95 spaces without a quality assessment suggested that most of these were not publically accessible and could therefore justifiably excluded from the analysis. PPG17 was replaced in 2012 by the National Planning Policy Framework (NPPF). PPG17 guidance set out the priorities and aimed for high standard open space provision, requiring local authorities to undertake an assessment of provision of open space, indoor facilities and outdoor sports. The assessment included site visits to greenspaces to rate a number of key criteria affecting quality.

The Sheffield PPG17 assessment criteria incorporate aspects of the Green Flag Programme, ILAM (Institute of Leisure and Amenity Management) Parks Management Guidance and the Tidy Britain Scheme. This resulted in a total of 68 measures collected for 19 different themes. A full breakdown of the measures collected can be found in Supplementary online Table S5. Each measure was scored on a scale of 1 to 5. The assessment considered a wide range of factors including accessibility, safety, management, maintenance, overall visual impression, and the presence of fixtures such as benches, bins, gates, signage, hedges, trees and paths. The assessment included a summary of all measures as an 'Overall Score'. In addition to the Overall Score, our work also used the Cleanliness theme that was generated from the following measures: Litter, Dog Fouling, Graffiti, and Chewing Gum.

Our analysis comparing the similarity between council field survey quality assessment and social media measures was undertaken for each greenspace (n=850). In contrast, data were aggregated to LSOAs (n=345) in order to explore the relationship between these variables and general health. Aggregation was achieved by using area weighting (Brindley et al., 2005) based on the quality weighted by the amount of greenspace within each LSOA. For example, a greenspace with twice the size would have twice the weighting in terms of its quality rating attributed to that LSOA. Of the 345 LSOAs in Sheffield, 32 contained no greenspace and so were excluded from the health related analysis.

2.4 Social media data

Our second aim was to determine whether social media data are reliable proxy measures of greenspace quality from field surveys. Measures extracted from social media included (1) the number of photographs taken within greenspaces and uploaded to Flickr (downloaded in August 2017 through the Flickr API); (2) the sentiment of those photograph captions; and (3) the number of geo-referenced Tweets from Twitter within greenspaces (extracted via the Twitter API for a one-year period from August 2016). Flickr (https://www.flickr.com) is a popular social networking platform allowing users to upload and share photographs. The photos include coordinates (longitude, latitude) showing where the images were taken. Twitter (https://twitter.com) is a microblogging service allowing users to share their messages known as "tweets". Only about 1% of tweets contain coordinate references (Crampton et al., 2013). Both georeferenced Flickr photos and Tweets can then be mapped within a GIS and the number of photos/tweets within greenspaces calculated. The number of Flickr images and Tweets were expressed relative to the size of the greenspace as the number per hectare. Whilst individually social media entries may represent individual impressions or current descriptions of the state of the greenspace, when taken together they *may* provide some insights collectively in terms of greenspace 'quality'.

The data, however, were highly skewed, with 504 greenspaces containing no Flickr photographs (53% of all of the greenspaces) and 715 without Tweets (75%). There are two interpretations of this. Firstly, these sites might have no social media data due to the nature of the sites themselves: people chose not to visit them due to their poor quality. In this interpretation, the absence of data represents poor quality (true zeros). Alternatively, data may be absent simply due to small numbers of social media users and thus be down to chance. According to this interpretation, some of the sites with no data may have good quality (excess zeros). To overcome this, further (sensitivity) analysis was undertaken which only included greenspaces where data were present (in other words – excluding those sites with no data).

In addition to the number of Flickr photos and Tweeters within greenspaces, our work also draws on the textual descriptions accompanying the Flickr data, which may provide insight into the 'quality' of the site (as previously shown in Fig. 1). We aggregated the text (title and description) for each Flickr photograph and undertook sentiment analysis using the Python module - VADER (Valence Aware Dictionary and sEntiment Reasoner) Sentiment Analysis – version 2.5. This is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media (Hutto and Gilbert, 2014), and has been used extensively in the literature in similar contexts (Anuta et al., 2017; Won et al., 2017). For the text of each photo, VADER generates a positive, neutral and negative sentiment score (whereby the three components sum to a value of 1). We required a measure of sentiment as a single value for each greenspace and so aggregated the values for each separate component for all the photos from each greenspace. This enabled us to calculate the ratio of positive to negative sentiment for every greenspace which we used as a further measure of potential greenspace quality.

As with the previously described greenspace data, individual - park-level data were used to investigate similarity to council field survey derived quality (aim 2 of our work). In order to explore the relationship with general health (aim 1), park-level data were aggregated to LSOAs using area weighting (as previously described in Section 2.3).

2.5 Socio-economic and demographic characteristics

We adjusted for area characteristics that were plausibly associated with general health. The income domain of the 2010 English index of multiple deprivation (EIMD) were used, being the closest time period to the 2011 health data. This domain has commonly been used to adjust for socio-economic deprivation (Dennis and James, 2017; Mitchell and Popham, 2008; Richardson and Mitchell, 2010).

Potential confounders of the total amount of greenspace and average garden size were also included within analysis (as used by Dennis and James, 2017; Brindley *et al.*, 2018). These measures were generated using detailed land cover data produced for Sheffield by Ersoy (2015), which consists of a composite of the Ordnance Survey (OS) MasterMap Topography layer, Land Cover Map 2007, Forestry Commission National Inventory Woodland and Trees Layer and Sheffield City Council Green and Open Spaces Layer.

Air pollution and smoking can potentially confound the association between greenspace and health because there might be less traffic pollution in areas of greater greenspace and higher levels of smoking in deprived areas which might have lower proportions of greenspace. Therefore, we also controlled for the levels of air pollution and smoking. Air pollution data consisted of 1km gridded estimates of Particulate Matter of ten microns in diameter or smaller (PM_{10}) modelled by the UK's Department for Environment, Food and Rural Affairs in the year 2010 and assigned to LSOAs by the population weighted average for each LSOA (where the population represented the census headcounts at unit postcode level). A proxy of smoking was obtained through new cases of lung cancer for the years 2010-2014, as a standardised incidence ratio, at Medium Layer Super Output Areas (MSOA) (available from www.localhealth.org.uk). Confounders for both smoking and air pollution have been commonly included in similar greenspace population level studies (Mitchell and Popham, 2008, 2007; Richardson et al., 2010; Richardson and Mitchell, 2010).

A variant for the smoking confounder was examined within sensitivity analysis to ensure robustness. The number of lung cancer related hospital admissions for the period 2002-2010 for each LSOA were analysed (as used by Brindley *et al.*, 2018). This measure replaced the smoking proxy (incidence of new cases of lung cancer at MSOA level) used within the main modelling. Whilst such data are available at a higher spatial resolution, they were only available tied to 2001 LSOA boundaries. For this reason, only LSOAs whose boundary did not change between the period 2001 to 2011 were included within analysis. This resulted in 20 LSOAs, with boundary changes, being excluded from this sensitivity analysis.

2.6 Analytical approach

Our analysis consisted of two separate stages. Firstly, we used Spearman correlation to explore the similarity between the field survey quality measures and social media derived proxies at the park level (aim 2 of our paper). A correlation matrix was generated between the quality measures derived from field surveys by the Council and those extracted from social media data.

Secondly, we tested whether there was an independent association between greenspace quality and self-reported poor health using negative binomial regression (aim 1). The analysis controlled for the confounding factors previously described (income deprivation, total greenspace, average garden size, smoking prevalence and levels of air pollution). The dependent variable was the total number of people reporting poor health, whilst the offset was the number expected given the age and sex

composition (indirect standardisation). Poisson models were rejected due to over dispersion. Analysis was undertaken within SAS version 9.4.

In total five separate models were run, as shown in Table 1, where the only difference was the independent variable relating to greenspace quality: council field survey for overall quality (Model 1); council field survey for cleanliness (Model 2); number of Flickr photos taken in greenspace (per hectare of greenspace) (Model 3); the sentiment of Flickr photos taken in greenspace (Model 4); and the number of unique Twitter users tweeting in greenspaces (per hectare) (Model 5).

The following sensitivity analysis was also undertaken to ensure robustness of findings. (1) The total number of Flickr photographs and Twitter users were used instead of measures that were relative to the size of the greenspace (such as photographs per hectare). It might be the case that the total number of users is a better measure for 'quality' rather than a measure in relation to size. (2) The social media derived quality proxies were replaced with modified variants whereby only sites where social media data were present were included in the analysis (hence ignoring those sites with zero Flickr photos or Tweets within the analysis). This was undertaken in order to ensure that findings were not influenced by the relatively large number of greenspaces where there were no social media data present (as previously described in Section 2.4). (3) The smoking prevalence variable was replaced with an alternative measure to ensure robustness. As previously described, this entailed using the number of lung cancer related hospital admissions for the period 2002-2010 at LSOA level (for LSOAs whose boundary remained unchanged between the period 2001-2011).

3. Results

According to the 2011 Census, the 345 LSOAs in Sheffield district included a total population of 552,698, with 34,372 residents (6.4%) reporting poor health. Within the 945 greenspaces, there were 29,660 Flickr photos (an average of 31 per greenspace) and 4,632 Tweets (average of 4.9 per greenspace) taken by 2,165 different (unique) Twitter users (average of 2.3 per greenspace). The distribution of social media data, however, was highly skewed (as previously discussed in Section 2.4). Only 443 of the 945 greenspaces had one or more Flickr photos and only 231 greenspaces contained a geo-located Tweet. For this reason, (as discussed within Section 2.4), analysis was run both for all greenspaces and also only for those containing social media data - in order to test the robustness of the findings. In contrast, there were five greenspaces containing over 1,000 photos (Park House Lane Sports Ground; Weston Park; Graves Park; Don Valley Bowl; Botanical Gardens) and five sites containing over 100 different Twitter users (Peace Gardens; Ponderosa; Endcliffe Park; Devonshire Green; Botanical Gardens).

Spearman correlation demonstrated a weak positive relationship between the overall site quality measure and the subcomponent related to cleanliness when comparing quality scores for each park (r=0.324, p <0.001).

3.1 Similarity between site survey quality measures and social media derived proxies (aim 2)

The correlation matrix between the different measures of quality is shown in Table 2. The levels of association between the Council field survey measures of quality (overall score and cleanliness) and those quality proxies derived from social media (Flickr and Twitter) appear generally poor. The highest correlation score was just 0.25 between the total number of (unique) Twitter users and the overall Council field survey measure. In contrast, there are generally encouraging levels of

9

agreement between the two sources of proxy quality measures extracted from the social media data (Flickr and Twitter). Comparable output was also found for the correlation matrix which excluded greenspaces where there were no social media data (see Supplementary online Table S1).

3.2 Relationship between greenspace quality and self-reported poor health (aim 1)

Our five models explored the relationship between greenspace quality and self-reported poor general health, after controlling for confounding factors, using a number of different variables to represent quality: (1) council field survey for overall quality; (2) council field survey for cleanliness; (3) number of Flickr photos taken in greenspace; (4) sentiment of Flickr photos taken in greenspace; and (5) number of unique Twitter users tweeting in greenspaces.

The independent relationship between the various measures of greenspace quality and self-reported poor general health are shown in Fig. 2 and Table 3. The main association observed was the relationship between cleanliness of greenspace and general health. We found that there was a higher prevalence ratio (PR) for populations residing in areas of lower cleanliness. The PR for poor health was 1.09 (95% CI 1.00-1.18) for the quintile with the lowest greenspace cleanliness compared against the quintile with the highest cleanliness, when accounting for income deprivation and confounders.

In contrast, there was no relationship with poor health for either the Twitter users or Flickr sentiment variables. Whilst there was a significant statistical relationship between poor general health and the numbers of unique Twitter users in greenspace (per hectare) (p = 0.03), there is no consistency across the quintiles. As such, any underlying relationship is unlikely – especially when the large confidence intervals are considered. Furthermore, although there might be some marginal trend with increasing poor health prevalence for both the Council field survey for overall quality and intensity of Flickr photos, neither are statistically significant – with large confidence intervals.

Within the sensitivity analysis, there was no association between poor health and either the number of Flickr photos or Twitter users when using the total, absolute numbers, rather than relative measures (per hectare of greenspace), as shown within Supplementary online Table S2. Similarly, as illustrated within Supplementary online Table S3, no association was present between poor health and the three social media derived measures of greenspace quality when those greenspaces without social media data were removed from analysis (in other words greenspaces with no social media data were excluded rather than representing a true zero). Finally, replacing the smoking proxy for the number of hospital admissions related to lung cancer did not alter the output (Supplementary online Table S4). The pattern of association remained similar between self-reported poor general health and greenspace quality as measured using the Council field survey of cleanliness.

When comparing the PR for poor health between the different variables within the model within Table 3, it is apparent that greenspace quality, in terms of cleanliness, is at least as important as the total amount of greenspace. This finding is considered further within the discussion.

4. Discussion

The purpose of this paper was twofold. Firstly, it contributes to filling a gap in our knowledge and understanding, as identified within the existing literature, concerning the role of greenspace quality

in the relationship between greenspace and health (aim 1). Secondly, it seeks to establish whether data extracted from social media might be a suitable replacement for costly field surveys concerning greenspace quality (aim 2).

4.1 The relationship between greenspace quality and self-reported health (aim 1)

The current literature has identified the need to go beyond simply measuring the amount of greenspace and to explore the role of greenspace 'quality' when exploring potential health benefits (Akpinar, 2016; Akpinar et al., 2016; De Vries et al., 2003; Maas et al., 2006; Markevych et al., 2017; Richardson et al., 2010; Richardson and Mitchell, 2010; van den Berg et al., 2007; Wyles et al., 2017).

Our results support our hypothesis that there is an association between poor general health and greenspace quality. This manifests itself in the association between poor general health and quality as measured by the Council cleanliness field survey. Our findings support those of previous studies (Agyemang et al., 2007; van Dillen et al., 2012) but at the same time also contrast with other studies which found no relationship with 'quality' (Putrik et al., 2015), or those that identified associations contra to expectation (Putrik et al., 2015; Wheeler et al., 2015). Based on previous findings we would suggest that cleanliness impacts on perceived general health by discouraging greenspace usage, thereby depriving potential users of the opportunity to experience the health and wellbeing benefits that green spaces provide (Lachowycz and Jones, 2013). Even if people do visit, the negative impact of encountering litter, dog fouling, graffiti and chewing gum may outweigh or detract from opportunities to benefits from restoration from stress and social interaction (Pretty et al., 2005; Wyles et al., 2016). Furthermore, greenspaces exhibiting low levels of cleanliness are likely to erode neighbourhood satisfaction and general wellbeing (Zhang et al., 2017). Our work highlights the need to analyse the relationship between greenspace and health in a more nuanced manner, incorporating elements of greenspace quality and in particular those relating to cleanliness.

Whilst a relationship was found between self-reported health and the Council cleanliness field survey, there was no significant association between health and the other measures of greenspace quality – either from the Council overall measure of quality or those derived from social media. Whilst a marginal trend might have existed whereby increasing quintiles of higher overall quality resulted in lower levels of poor health (see Fig. 2), the relationship was not statistically significant and was weaker than the association between cleanliness and poor health (both within the main modelling and sensitivity analysis). The overall quality field survey by the Council centred on a wide range of 68 'quality' measures across 19 themes. As such it included scoring of a wide range of factors such as clear signage for parking, adequate mowing margins and well defined pathway edges. Not all of these measures may be indicators of what greenspace 'quality' means to users and further research is therefore required to unpack the dynamics of what elements of 'quality' may contribute to health benefits.

4.2 The potential for social media to proxy field survey measures of greenspace quality (aim 2)

In examining the potential for social media data to proxy field surveys for greenspace quality, we found that data extracted from Flickr and Twitter did not appear to be similar to measures derived from Council field surveys for overall quality or cleanliness. This does not necessarily mean that they may not relate to other elements of quality or have wider possible applications. For example, such data may provide indications of likely usage of greenspaces.

The lack of correlation between measures of quality based on primary field surveys of local environments and social media based assessments may be because (i) increased data coverage from social media are required for a more robust analysis; (ii) social media may be addressing different aspects of quality; or (iii) they are not good indicators of greenspace quality.

The level of similarity between measures from the two different sources of social media (Flickr and Twitter), however, was encouraging: with a correlation of 0.51 between the number of Flickr photographs and number of unique Twitter users. Further discussion of the limitations concerning the social media data can be found in Section 4.3.

4.3 Strengths and weaknesses

Our population level study used the power of secondary datasets to investigate the relationship between greenspace quality and poor self-reported general health. The study built upon the existing literature and is hypothesis driven. One of the strengths of our study is that it addresses a gap within the existing literature as suggested by numerous recent studies (Akpinar, 2016; Akpinar et al., 2016; De Vries et al., 2003; Maas et al., 2006; Richardson et al., 2010; Richardson and Mitchell, 2010; van den Berg et al., 2007; Wyles et al., 2017). A further strength of our paper is the innovative use of data extracted from social media to potentially proxy greenspace quality as measured using field surveys. Such field surveys can be costly, time-consuming and impractical to collect for numerous greenspaces across extensive geographic areas. Whilst alternative measures of auditing not requiring park visitation such as Public Open Space Desktop Auditing Tool (POSDAT) have demonstrated great potential (Edwards et al., 2013) they are limited in their detection of the extent of littering and graffiti to those areas of greenspace that are visible within Google Street View imagery.

As with any population level, small-area study there are a number of limitations. Firstly, it is a misconception to believe that correlation implies causation. Whilst our study reported an association between greenspace cleanliness and self-reported health, we cannot confirm whether there is a causal link. Other risk factors, not controlled for within our study, may act as a third factor within the association between greenspace quality and health. We acknowledge that additional variables such as walkability, traffic noise and population density may also affect general health and need to be considered in future studies. Further work, including cohort studies, would be required in order to comprehensively test for causation.

Secondly, the associations that we have found are based on data aggregated to bounded administrative units (in this case LSOAs). As such, it should not be assumed that the same associations would hold at the individual level. This is a well-studied phenomenon, termed the ecological fallacy (Haining, 2003). Our analysis did not incorporate adjustment for spatially structured random effects. We have, however, adjusted for a number of potential confounders which should have minimised the effect of residual spatial autocorrelation.

There are also a number of limitations that are more specific to the precise nature of our study. Firstly, we recognise that we treat greenspace as a single class of object instead of breaking it down by categories of type (for example distinguishing between areas of woodland, amenity greenspace and cemeteries and so forth). Whilst we completely advocate the need to not treat greenspace as homogenous spaces, for practicality reasons it would be impossible to break down our measures by greenspace types, as this would result in too many of our statistical units (LSOAs) having no data for particular classes of greenspace; so that we would lose too much statistical power.

12

This corresponds with our second limitation, that our study, being tied to Sheffield, has a rather limited lack of power, leading to relatively large confidence intervals within the work. Ideally, the analysis would be run for a wider geographical extent to help resolve this. The issue here, however, is that we are dependent on the requirement for adequate field survey data measuring greenspace quality, which only exist on an ad hoc basis. This was precisely why we undertook the second stage of our work to test whether alternative proxies of quality might be found that facilitated the collection of greenspace quality measures without the need for costly and time-consuming field surveys.

As a population level study, our measure of greenspace relates to provision and has not taken into account the extent to which use of the greenspace occurs. Complex patterns are likely to exist between use of greenspace and home locations – with residents not always using greenspaces in their local geographic area. It is possible that our lung cancer proxy for smoking is also affected by air pollution. Results (not shown), however, were consistent when excluding the lung cancer proxy. Furthermore, we recognise that the level of cleanliness in any greenspace is likely to change over time.

Finally, whilst it is clear that social media data has the potential to provide insight and deeper understanding into the use and values of greenspace (see Fig. 1 and previous work such as described within Section 1.3), such an approach is not without its issues and limitations. We acknowledge the many issues that frequently accompany the use of social media data – including the bias that exists within their usage (Ruths and Pfeffer, 2014) – which will also affect their geographic coverage. The highly skewed nature of social media data, with a small number of greenspaces contributing for the majority of content, raises concerns of representativeness. For example, the high number of photos associated with certain greenspaces (such as Park House Lane Sports Ground and Don Valley Bowl) were due to sporting events. Given that the intensity of social media content has been found to be positively correlated with sociodemographic factors including income, youth, and education (Li et al., 2013), it is possible that social media users within greenspace may not be representative of greenspace users as whole. Such effects may potentially bias results and need to be taken into consideration when interpreting results from the use of such data as proxies for greenspace quality.

5. Conclusions and recommendations

Whilst our findings showed that measures extracted from social media appear to make poor proxies for greenspace quality, we were also able to demonstrate the association between greenspace quality (in terms of cleanliness) and self-reported general health. This relationship persists even when accounting for a range of potential confounders. Whilst affected by the limitations associated with any population level design, our paper contributes to the evidence base for this topic and demonstrates the need for further research in this area.

People living in areas with poorer quality of greenspace have higher levels of poor health. Even though causality needs confirmation, our work highlights the potential importance of greenspace cleanliness. The health and wellbeing pathways provided by greenspaces are redundant if no-one wants to visit them. Our findings have particular importance to greenspace management. Greenspace cleanliness appears to be an important factor to help alleviate poor general health.

Acknowledgements

Conflicts of interest

The authors have no conflicts of interest to declare.

References

Agyemang, C., Hooijdonk, C. v., Wendel-Vos, W., Lindeman, E., Stronks, K., Droomers, M., 2007. The association of neighbourhood psychosocial stressors and self-rated health in Amsterdam, The Netherlands. J. Epidemiol. Community Heal. 61, 1042–1049. https://doi.org/10.1136/jech.2006.052548

Akpinar, A., 2016. How is quality of urban green spaces associated with physical activity and health? Urban For. Urban Green. 16, 76–83. https://doi.org/10.1016/j.ufug.2016.01.011

Akpinar, A., Barbosa-Leiker, C., Brooks, K.R., 2016. Does green space matter? Exploring relationships between green space type and health indicators. Urban For. Urban Green. 20, 407–418. https://doi.org/10.1016/j.ufug.2016.10.013

Anuta, D., Churchin, J., Luo, J., 2017. Election Bias : Comparing Polls and Twitter in the 2016 U .S. Election. CoRR abs/1701.0, arXiv preprint arXiv:1701.06232.

Barker, A., Booth, N., Churchill, D., Crawford, A., 2017. The Future Prospects of Urban Public Parks. Report available at: https://futureofparks.leeds.ac.uk/wp-content/uploads/sites/26/2017/07/Job-38853-Future-of-Parks-Findings-Report.pdf [last accessed 23/01/2019]

Bedimo-Rung, A.L., Mowen, A.J., Cohen, D.A., 2005. The Significance of Parks to Physical Activity and. Am. J. Prev. Med. 28, 159–168. https://doi.org/10.1016/j.ampre.2004.10.024

Bosch, M. Van Den, Sang, Å.O., 2017. Urban natural environments as nature-based solutions for improved public health – A systematic review of reviews. Environ. Res. 158, 373–384. https://doi.org/10.1016/j.envres.2017.05.040

Bowler, D., Buyung-Ali, L., Knight, T., Pullin, A., 2010. A systematic review of evidence for the added benefits to health of exposure to natural environments. BMC Public Health 10, 456. https://doi.org/10.1186/1471-2458-10-456

Brindley, P., Jorgensen, A., Maheswaran, R., 2018. Domestic gardens and self‐reported health : a national population study. Int. J. Health Geogr. 17. https://doi.org/10.1186/s12942-018-0148-6

Brindley, P., Wise, S.M., Maheswaran, R., Haining, R.P., 2005. The effect of alternative representations of population location on the areal interpolation of air pollution exposure. Comput. Environ. Urban Syst. 29, 455–469. https://doi.org/10.1016/j.compenvurbsys.2004.01.001

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Wilson, M.W., Zook, M., Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., 2013. Beyond the geotag : situating ' big data ' and

leveraging the potential of the geoweb. Cartogr. Geogr. Inf. Sci. 40, 130–139. https://doi.org/10.1080/15230406.2013.777137

Dadvand, P., Bartoll, X., Basagaña, X., Dalmau-bueno, A., Martinez, D., Ambros, A., Cirach, M., Triguero-mas, M., Gasconab, M., Borrell, C., Nieuwenhuijsen, M.J., 2016. Green spaces and General Health : Roles of mental health status , social support , and physical activity. Environ. Int. 91, 161–167. https://doi.org/10.1016/j.envint.2016.02.029

De Vries, S., van Dillen, S.M.E., Groenewegen, P.P., Spreeuwenberg, P., 2013. Streetscape greenery and health: Stress, social cohesion and physical activity as mediators. Soc. Sci. Med. 94, 26–33. https://doi.org/10.1016/j.socscimed.2013.06.030

De Vries, S., Verheij, R.A., Groenewegen, P.P., Spreeuwenberg, P., 2003. Natural environments - healthy environments? An exploratory analysis of the relationship between green space and health. Environ. Plan. A 35, 1717–1731. https://doi.org/10.1068/a35111

Dennis, M., James, P., 2017. Evaluating the relative influence on population health of domestic gardens and green space along a rural-urban gradient. Landsc. Urban Plan. 157, 343–351. https://doi.org/10.1016/j.landurbplan.2016.08.009

Department for Work and Pensions, 2014. The Use of Social Media for Research and Analysis : A Feasibility Study. Report avilable at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_dat a/file/387591/use-of-social-media-for-research-and-analysis.pdf [last accessed 23/01/2019]

Desalvo, K.B., Bloser, N., Reynolds, K., He, J., Muntner, P., 2006. Mortality Prediction with a Single General Self-Rated Health Question. J. Gen. Intern. Med. 21, 267–275. https://doi.org/10.1111/j.1525-1497.2005.0291.x

Donahue, M.L., Keeler, B.L., Wood, S.A., Fisher, D.M., Hamstead, Z.A., 2018. Landscape and Urban Planning Using social media to understand drivers of urban park visitation in the Twin Cities , MN. Landsc. Urban Plan. 175, 1–10. https://doi.org/10.1016/j.landurbplan.2018.02.006

Dunkel, A., 2015. Landscape and Urban Planning Visualizing the perceived environment using crowdsourced photo geodata. Landsc. Urban Plan. 142, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

Dzhambov, A.M., Dimitrova, D.D., Dimitrakova, E.D., 2014. Urban Forestry & Urban Greening Association between residential greenness and birth weight : Systematic review and meta-analysis. Urban For. Urban Green. 13, 621–629. https://doi.org/10.1016/j.ufug.2014.09.004

Edwards, N., Hooper, P., Trapp, G.S.A., Bull, F., Boruff, B., Giles-Corti, B., 2013. Development of a Public Open Space Desktop Auditing Tool (POSDAT): A remote sensing approach. Appl. Geogr. 38, 22–30. https://doi.org/10.1016/j.apgeog.2012.11.010

Figueroa-Alfaro, R.W., Tang, Z., 2017. Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. J. Environ. Plan. Manag. 60, 266–281. https://doi.org/10.1080/09640568.2016.1151772

Gascon, M., Triguero-mas, M., Martínez, D., Dadvand, P., 2016. Residential green spaces and mortality : A systematic review. Environ. Int. 86, 60–67. https://doi.org/10.1016/j.envint.2015.10.013

Gascon, M., Triguero-mas, M., Martínez, D., Dadvand, P., 2015. Mental Health Benefits of Long-Term Exposure to Residential Green and Blue Spaces : A Systematic Review. Int. J. Environ. Res. Public Health 12, 4354–4379. https://doi.org/10.3390/ijerph120404354

Gerrard, D., Sykora, M., Jackson, T., 2017. Social media analytics in museums: extracting expressions of inspiration. Museum Manag. Curatorsh. 32, 232–250. https://doi.org/10.1080/09647775.2017.1302815

Giles-Corti, B., Broomhall, M.H., Knuiman, M., Collins, C., Douglas, K., Ng, K., Lange, A., Donovan, R.J., 2005. Increasing walking: How important is distance to, attractiveness, and size of public open space? Am. J. Prev. Med. 28, 169–176. https://doi.org/10.1016/j.amepre.2004.10.018

Haining, R., 2003. Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511754944

Hartig, T., Mitchell, R., de Vries, S., Frumkin, H., 2014. Nature and Health. Annu. Rev. Public Health 35, 207–228. https://doi.org/10.1146/annurev-publhealth-032013-182443

Hillsdon, M., Panter, J., Foster, C., Jones, A., 2006. The relationship between access and quality of urban green space with population physical activity. Public Health 120, 1127–1132. https://doi.org/10.1016/j.puhe.2006.10.007

Hutto, C.J., Gilbert, E., 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14). June 2014, Ann Arbor; MI, pp. 216–225.

Idler, E.L., Benyamini, Y., 1997. Self-Rated Health and Mortality : A Review of Twenty-Seven Community Studies. J. Health Soc. Behav. 38, 21–37. https://www.jstor.org/stable/2955359

James, P., Banay, R.F., Hart, J.E., Laden, F., 2015. A Review of the Health Benefits of Greenness. Curr. Epidemiol. Reports 2, 131–142. https://doi.org/10.1007/s40471-015-0043-7

Keijzer, C. De, Gascon, M., Nieuwenhuijsen, M.J., 2016. Long-Term Green Space Exposure and Cognition Across the Life Course : a Systematic Review. Curr. Environ. Heal. Reports 3, 468–477. https://doi.org/10.1007/s40572-016-0116-x

Kyffin, R.G., Goldacre, M.J., Gill, M., 2004. Mortality rates and self reported health: database analysis by English local authority area. Br. Med. J. 329, 887–888. https://doi.org/10.1136/bmj.38238.508021.F7

Lachowycz, K., Jones, A.P., 2013. Landscape and Urban Planning Towards a better understanding of the relationship between greenspace and health : Development of a theoretical framework. Landsc. Urban Plan. 118, 62–69. https://doi.org/10.1016/j.landurbplan.2012.10.012

Lanki, T., Siponen, T., Ojala, A., Korpela, K., Pennanen, A., Tiittanen, P., Tsunetsugu, Y., Kagawa, T., Tyrväinen, L., 2017. Acute effects of visits to urban green environments on cardiovascular physiology in women : A field experiment. Environ. Res. 159, 176–185. https://doi.org/10.1016/j.envres.2017.07.039

Lee, A.C.K., Maheswaran, R., 2011. The health benefits of urban green spaces: A review of the evidence. J. Public Health (Bangkok). 33, 212–222. https://doi.org/10.1093/pubmed/fdq068

Lovell, R., 2016. Natural England Access to Evidence Notes EIN017-021.Reports avilable at: http://publications.naturalengland.org.uk/category/6159558569361408 [last accessed 23/01/2019]

Maas, J., Verhij, R.., Groenewegen, P., De Vries, S., Spreeuwenberg, P., 2006. Green space, urbanity, and health: how strong is the relation? J. Epidemiol. Community Heal. 60, 587–592. https://doi.org/10.1136/jech.2005.043125

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A.M., Vries, S. De, Triguero-mas, M., Brauer, M., Nieuwenhuijsen, M.J., Lupp, G., Richardson, E.A., Astell-burt, T.,

Dimitrova, D., Feng, X., Sadeh, M., Standl, M., Heinrich, J., Fuertes, E., 2017. Exploring pathways linking greenspace to health : Theoretical and methodological guidance. Environ. Res. 158, 301–317. https://doi.org/10.1016/j.envres.2017.06.028

Maruthaveeran, S., Konijnendijk, C.C., Bosch, V. Den, 2014. Urban Forestry & Urban Greening A socio-ecological exploration of fear of crime in urban green spaces – A systematic review. Urban For. Urban Green. 13, 1–18. https://doi.org/10.1016/j.ufug.2013.11.006

Mavaddat, N., Kinmonth, A.L., Sanderson, S., Surtees, P., Bingham, S., Khaw, K.T., 2011. What determines self-rated health (SRH)? A cross-sectional study of SF-36 health domains in the EPIC-Norfolk cohort. J. Epidemiol. Community Health 65, 800–806. https://doi.org/10.1136/jech.2009.090845

Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. Lancet 372, 1655–60. https://doi.org/10.1016/S0140-6736(08)61689-X

Mitchell, R., Popham, F., 2007. Greenspace, urbanity and health: relationships in England. J. Epidemiol. Community Health 61, 681–3. https://doi.org/10.1136/jech.2006.053553

Ord, K., Mitchell, R., Pearce, J., 2013. Is level of neighbourhood green space associated with physical activity in green space? Int. J. Behav. Nutr. Phys. Act. 10, 127. https://doi.org/10.1186/1479-5868-10-127

Pretty, J., Peacock, J., Sellens, M., Griffin, M., 2005. The mental and physical health outcomes of green exercise. Int. J. Environ. Health Res. 15, 319–337. https://doi.org/10.1080/09603120500155963

Procter, R., Vis, F., Voss, A., 2013. Reading the riots on Twitter: Methodological innovation for the analysis of big data. Int. J. Soc. Res. Methodol. 16, 197–214. https://doi.org/10.1080/13645579.2013.774172

Putrik, P., de Vries, N.K., Mujakovic, S., van Amelsvoort, L., Kant, Ij.J., Kunst, A.E., van Oers, H., Jansen, M., 2015. Living Environment Matters: Relationships Between Neighborhood Characteristics and Health of the Residents in a Dutch Municipality. J. Community Health 40, 47–56. https://doi.org/10.1007/s10900-014-9894-y

Richardson, E., Mitchell, R., 2010. Gender differences in relationships between urban green space and health in the United Kingdom. Soc. Sci. Med. 71, 568–75. https://doi.org/10.1016/j.socscimed.2010.04.015

Richardson, E., Pearce, J., Mitchell, R., Day, P., Kingham, S., 2010. The association between green space and cause-specific mortality in urban New Zealand: an ecological analysis of green space utility. BMC Public Health 10, 240. https://doi.org/10.1186/1471-2458-10-240

Sessions, C., Wood, S.A., Rabotyagov, S., Fisher, D.M., 2016. Measuring recreational visitation at U . S . National Parks with crowd- sourced photographs. J. Environ. Manage. 183, 703–711. https://doi.org/10.1016/j.jenvman.2016.09.018

Short, M.E., Goetzel, R.Z., Pei, X., Tabrizi, M.J., Ozminkowski, R.J., Gibson, T.B., Dejoy, D.M., Wilson, M.G., 2009. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. J Occup Env. Med 51, 786–796. https://doi.org/10.1016/j.jacc.2007.01.076.White

Stark, J.H., Neckerman, K., Lovasi, G.S., Quinn, J., Weiss, C.C., Bader, M.D.M., Konty, K., Harris, T.G., Rundle, A., 2014. The impact of neighborhood park access and quality on body mass index among adults in New York City ☆. Prev. Med. (Baltim). 64, 63–68.

https://doi.org/10.1016/j.ypmed.2014.03.026

Tsakalidis, A., Liakata, M., Damoulas, T., Jellinek, B., Guo, W., Cristea, A., 2016. Combining Heterogeneous User Generated Data to Sense Well-being. Proc. COLING 2016, 26th Int. Conf. Comput. Linguist. Tech. Pap. 3007–3018.

van den Berg, A.E., Hartig, T., Staats, H., 2007. Preference for nature in urbanized societies: Stress, restoration, and the pursuit of sustainability. J. Soc. Issues 63, 79–96. https://doi.org/10.1111/j.1540-4560.2007.00497.x

Van Den Berg, M., Wendel-Vos, W., Van Poppel, M., Kemper, H., Van Mechelen, W., Maas, J., 2015. Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. Urban For. Urban Green. 14, 806–816. https://doi.org/10.1016/j.ufug.2015.07.008

van Dillen, S.M.E., de Vries, S., Groenewegen, P.P., Spreeuwenberg, P., 2012. Greenspace in urban neighbourhoods and residents' health: adding quality to quantity. J. Epidemiol. Community Health 66, e8–e8. https://doi.org/10.1136/jech.2009.104695

Vujcic, M., Tomicevic-dubljevic, J., Zivojinovic, I., Toskovic, O., 2018. Connection between urban green areas and visitors' physical and mental well-being. Urban For. Urban Green. https://doi.org/10.1016/j.ufug.2018.01.028

Wheeler, B.W., Lovell, R., Higgins, S.L., White, M.P., Alcock, I., Osborne, N.J., Husk, K., Sabel, C.E., Depledge, M.H., 2015. Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. Int. J. Health Geogr. 14, 17. https://doi.org/10.1186/s12942-015-0009-5

Won, D., Steinert-Threlkeld, Z.C., Joo, J., 2017. Protest Activity Detection and Perceived Violence Estimation from Social Media Images. CoRR abs/1709.0, arXiv preprint arXiv:1709.06204.

Wyles, K.J., Pahl, S., Thomas, K., Thompson, R.C., 2016. Factors That Can Undermine the Psychological Benefits of Coastal Environments. Environ. Behav. 48, 1095–1126. https://doi.org/10.1177/0013916515592177

Wyles, K.J., White, M.P., Hattam, C., Pahl, S., King, H., Austen, M., 2017. Are Some Natural Environments More Psychologically Beneficial Than Others? The Importance of Type and Quality on Connectedness to Nature and Psychological Restoration. Environ. Behav. 001391651773831. https://doi.org/10.1177/0013916517738312

Zanten, B.T. Van, Berkel, D.B. Van, Meentemeyer, R.K., Smith, J.W., Tieskens, K.F., 2016. Continental-scale quantification of landscape values using social media data. Proc. Natl. Acad. Sci. United States Am. 113, 12974–12979. https://doi.org/10.1073/pnas.1614158113

Zhang, Y., Berg, A.E. Van Den, Dijk, T. Van, Weitkamp, G., 2017. Quality over Quantity : Contribution of Urban Green Space to Neighborhood Satisfaction. Int. J. Environ. Res. Public Health 14, 535. https://doi.org/10.3390/ijerph14050535

Figures

Fig 1: Word clouds of the words used in the captions and titles of Flickr photographs taken within two contrasting greenspaces in Sheffield (size of the word reflects the number of occurrences)

a) greenspace with positive sentiment   b) greenspace with negative sentiment

Fig 2: Prevalence ratios (adjusted for income deprivation, total greenspace, average garden size, air pollution (PM_{10}), smoking prevalence) for poor general health in relation to each of the five measures of greenspace quality (with 95% CI)
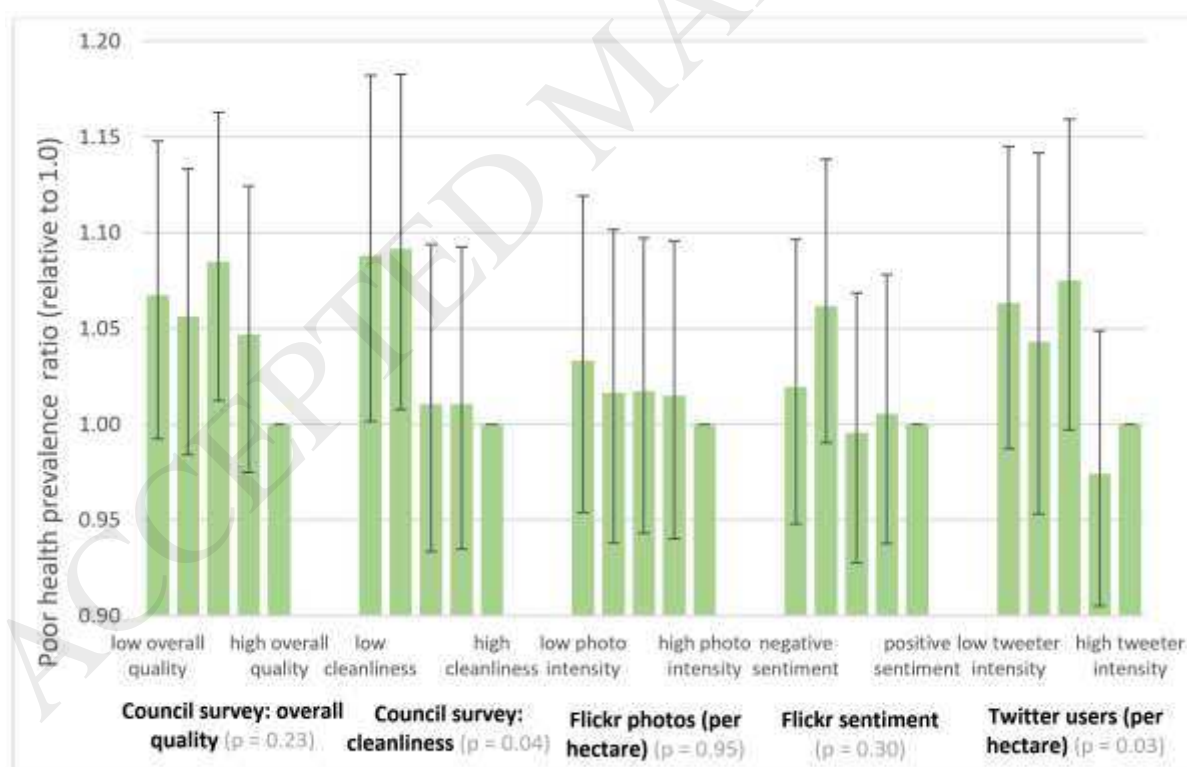
Tables

Table 1: Breakdown of the variables of the five models used to explore the relationship between greenspace quality of self-reported poor general health

	Model One	**Model Two**	**Model Three**	**Model Four**	**Model Five**
Dependent variable:	Self-reported poor general health				
Confounding factors:	Income deprivation Total greenspace Average garden size Air pollution levels (PM10) Smoking prevalence				
Greenspace quality measures:	Council field survey for overall quality	Council field survey for cleanliness	Flickr photos per hectare	Flickr sentiment	(Unique) Twitter users per hectare

Table 2: The level of correlation between the different measures of greenspace 'quality' (at the individual park level):

	PPG17: overall score	PPG17: cleanliness score	Number of Flickr photos	Flickr photos per hectare	Flickr sentiment (pos / neg)	Total (unique) Twitter users	Total Twitter users per hectare	Total Tweets per hectare
PPG17 – overall score	-	0.34	0.23	0.23	0.11	0.25	0.24	0.24
PPG17 – cleanliness score		-	0.06	0.08	0.03	0.08	0.08	0.08
Number of Flickr photos			-	0.96	0.45	0.51	0.47	0.47
Flickr photos per hectare				-	0.39	0.44	0.42	0.42
Flickr sentiment (pos / neg)					-	0.32	0.29	0.29
Total (unique) Twitter users						-	0.99	0.97
Total Twitter users per hectare							-	0.98
Total Tweets per hectare								-

Table 3: Regression results at the LSOA level: association between poor general health and modelled output

Variable:	Quintile:	Prevalence ratio	95% CI
Income deprivation	1 (most deprived)	2.76	2.50, 3.05
	2	2.28	2.08, 2.50
	3	1.68	1.54, 1.83
	4	1.26	1.15, 1.36
	5 (least deprived)	1.00	
Total greenspace	1 (least greenspace)	1.08	0.99, 1.18
	2	1.02	0.95, 1.09
	3	1.02	0.95, 1.09
	4	1.00	0.94, 1.06
	5 (most greenspace)	1.00	
Average garden size	1 (small gardens)	1.13	1.03, 1.24
	2	1.04	0.95, 1.13
	3	1.00	0.92, 1.09
	4	1.04	0.96, 1.12
	5 (large gardens)	1.00	
Air pollution (PM10)	1 (high air pollution)	1.05	0.97, 1.14
	2	1.07	0.99, 1.16
	3	0.99	0.92, 1.08
	4	1.01	0.93, 1.08
	5 (low air pollution)	1.00	
Smoking prevalence proxy	1 (high smoking levels)	1.22	1.11, 1.34
	2	1.16	1.06, 1.27
	3	1.14	1.05, 1.24
	4	1.06	0.97, 1.14
	5 (low smoking levels)	1.00	
The following variables were introduced one at a time alongside the variables above:			
Model 1: Council field survey - overall quality	1 (low overall quality)	1.07	0.99, 1.15
	2	1.06	0.98, 1.13
	3	1.08	1.01, 1.16
	4	1.05	0.97, 1.12
	5 (high overall quality)	1.00	
Model 2: Council field survey - cleanliness	1 (low cleanliness)	1.09	1.00, 1.18
	2	1.09	1.01, 1.18
	3	1.01	0.93, 1.09
	4	1.01	0.93, 1.09
	5 (high cleanliness)	1.00	
Model 3: Flickr photos per hectare	1 (low number of photos)	1.03	0.95, 1.12
	2	1.02	0.94, 1.10
	3	1.02	0.94, 1.10
	4	1.01	0.94, 1.10
	5 (high number of photos)	1.00	
Model 4: Flickr sentiment	1 (negative sentiment)	1.02	0.95, 1.10
	2	1.06	0.99, 1.14
	3	1.00	0.93, 1.07
	4	1.01	0.94, 1.08
	5 (positive sentiment)	1.00	
Model 5: Number of Twitter users per hectare	1 (low number of tweeters)	1.06	0.99, 1.14
	2	1.04	0.95, 1.14
	3	1.07	1.00, 1.16
	4	0.97	0.91, 1.05
	5 (high number of tweeters)	1.00	