

BMJ Open Longitudinal access and exposure to green-blue spaces and individual-level mental health and well-being: protocol for a longitudinal, population-wide record-linked natural experiment

Amy Mizen,¹ Jiao Song,¹ Richard Fry,¹ Ashley Akbari,¹ Damon Berridge,¹ Sarah C Parker,¹ Rhodri Johnson,¹ Rebecca Lovell,² Ronan A Lyons,¹ Mark Nieuwenhuijsen,³ Gareth Stratton,⁴ Benedict W Wheeler,² James White,⁵ Mathew White,² Sarah E Rodgers^{1,6}

To cite: Mizen A, Song J, Fry R, *et al.* Longitudinal access and exposure to green-blue spaces and individual-level mental health and well-being: protocol for a longitudinal, population-wide record-linked natural experiment. *BMJ Open* 2019;**9**:e027289. doi:10.1136/bmjopen-2018-027289

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-027289>).

AM and JS contributed equally.

Received 15 October 2018

Revised 25 October 2018

Accepted 31 October 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Professor Sarah E Rodgers; Sarah.Rodgers@liverpool.ac.uk

ABSTRACT

Introduction Studies suggest that access and exposure to green-blue spaces (GBS) have beneficial impacts on mental health. However, the evidence base is limited with respect to longitudinal studies. The main aim of this longitudinal, population-wide, record-linked natural experiment, is to model the daily lived experience by linking GBS accessibility indices, residential GBS exposure and health data; to enable quantification of the impact of GBS on well-being and common mental health disorders, for a national population.

Methods and analysis This research will estimate the impact of neighbourhood GBS access, GBS exposure and visits to GBS on the risk of common mental health conditions and the opportunity for promoting subjective well-being (SWB); both key priorities for public health. We will use a Geographic Information System (GIS) to create quarterly household GBS accessibility indices and GBS exposure using digital map and satellite data for 1.4 million homes in Wales, UK (2008–2018). We will link the GBS accessibility indices and GBS exposures to individual-level mental health outcomes for 1.7 million people with general practitioner (GP) data and data from the National Survey for Wales (n=~12 000) on well-being in the Secure Anonymised Information Linkage (SAIL) Databank. We will examine if these associations are modified by multiple sociophysical variables, migration and socioeconomic disadvantage. Subgroup analyses will examine associations by different types of GBS. This longitudinal study will be augmented by cross-sectional research using survey data on self-reported visits to GBS and SWB.

Ethics and dissemination All data will be anonymised and linked within the privacy protecting SAIL Databank. We will be using anonymised data and therefore we are exempt from National Research Ethics Committee (NREC). An Information Governance Review Panel (IGRP) application (Project ID: 0562) to link these data has been approved. The research programme will be undertaken in close collaboration with public/patient involvement groups. A multistrategy programme of dissemination is planned

Strengths and limitations of this study

- This retrospective controlled evaluation of a natural experiment includes the majority of adults in Wales, UK between 2008 and 2018, minimising selection bias.
- Generating a national longitudinal dataset of changes in green-blue spaces (GBS) exposure and access to GBS for households will reduce ecological fallacy.
- Spatial and temporal accessibility data linked for individuals and their routinely captured health service use, together with potential confounders, will allow us to investigate the impact of GBS on well-being.
- Detailed self-report data on well-being and GBS visit behaviour from cross-sectional national survey data will be linked to health service utilisation data.
- Despite including a large number of potentially confounding variables, this non-randomised study using routinely recorded data may omit some unknown confounders, thereby introducing a moderate level of bias due to confounding.

with the academic community, policy-makers, practitioners and the public.

INTRODUCTION

Globally, 686 million people suffer with common mental health disorders (CMDs) such as depression or anxiety.¹ In the UK, CMDs are experienced by around one in four of the population, and mental ill health costs the economy over £100 billion per annum in health, social care and quality of life loss costs.^{2–3} Subjective well-being (SWB) is also related to mental and physical health outcomes, including life expectancy,^{4–5} and is a key marker of quality of life.⁶ With increasing impacts on wider societal costs,

CMDs and promoting SWB are growing in importance. Access to natural environments—considered here as ‘green-blue spaces’ (GBS)—such as parks and beaches—may provide opportunities to support and promote good public mental health and well-being.^{7,8} The evidence base on the impacts of GBS, on mental health and well-being is growing rapidly. Current research suggests that the benefits may differ by population group, context and health outcome.^{9,10}

Recent systematic reviews, of predominantly cross-sectional studies, indicate positive relationships between mental health and SWB outcomes with living near GBS.^{11,12} Access to GBS is positively associated with GBS use, and using GBS may improve health outcomes through a number of mechanisms. For example, increased physical activity,^{13,14} psychological restoration,^{15,16} noise mitigation,¹⁷ heat and humidity regulation,^{18,19} increased social interaction and cohesion.^{20–22} Contrary to these findings, a study in New Zealand found no evidence that GBS influenced cardiovascular disease mortality²³ and suggested that GBS and health relationships may vary according to country or environmental contexts.

Many studies have found that access to GBS can vary across socioeconomic status (SES) areas; with more deprived areas tending to have poorer access to GBS.^{24,25} Differences in the distribution of GBS across SES may influence and contribute to SES-related health inequalities.²⁶ Furthermore, individuals from higher SES groups are more likely to select living in greener neighbourhoods.²⁷ Previous studies lack censoring for births, deaths and migration, which may have led to health selection effects which can alter the strength of association. There is also a lack of large-scale population-level studies which have systematically and explicitly examined associations within and between subgroups other than SES¹⁰ (eg, by gender,^{28–30} age, education^{22,31}). Furthermore, ethnic minority groups are under-represented in studies due

to selective non-response.³² Studies that have focused on differences between ethnic minorities found positive effects of GBS on well-being differed by ethnicity³³ and ethnic minorities had poorer access to GBS.³⁴

Study methods used to define access to GBS differ; some studies have calculated small area-level metrics and others have used distances. There is no accepted method to define GBS or for measuring different types of GBS (eg, according to general land cover categories, ecosystems or landscape classifications).^{35–41} However, it is difficult to expect an authoritative voice on defining GBS access because how people engage with GBS varies by population subgroup and environmental context. Therefore, how GBS is defined will vary by study design and the focus of the research. We propose to take an approach that will consider land use but also access points, rights of way and amenities such as benches. There are also differences in study aims because of differing perspectives; such as health research, environmental research and policy. Few studies have also considered the issue of access to blue spaces,^{42,43} which has also contributed to diversity in GBS accessibility measures. Although there is now a significant body of research, there are few longitudinal studies on mental health outcomes in response to changes in GBS access or with assessments of actual visits to GBS.^{44–47}

We define GBS accessibility indices as metrics that represent how people may access GBS (figure 1). GBS exposure is the ambient environment that people experience because of where they live and will be defined as greenness immediately surrounding home/can be seen out of the window. By including both of these measures in our study, we will be able to build a more holistic view of how GBS impacts CMD and well-being.

Aim

Our primary aim is to quantify the impact of changes in access to GBS and ambient GBS exposure on CMD and

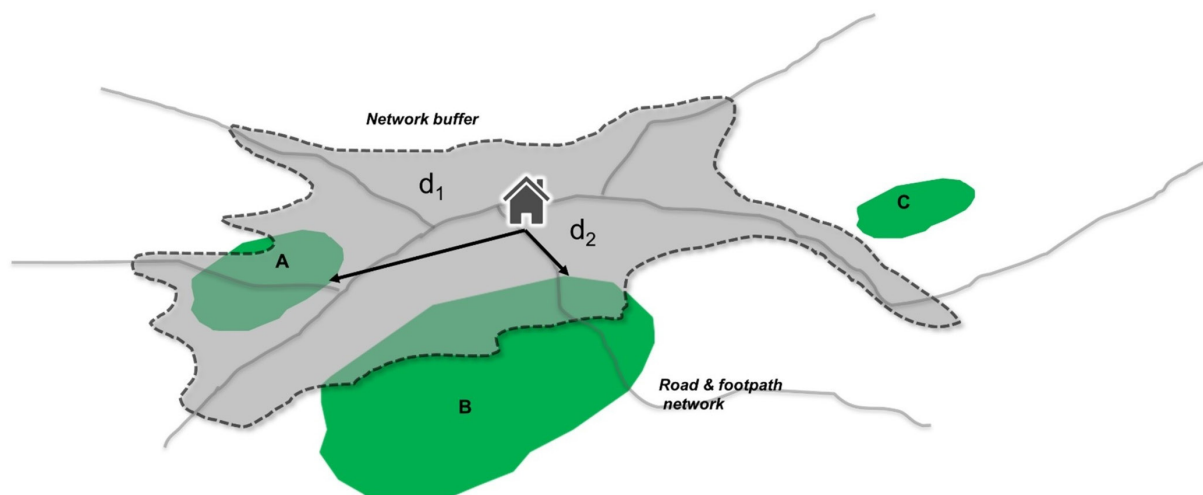


Figure 1 Residential GBS access for each home. (1) Network buffer defines household access boundary (eg, 10 min walk). (2) GBS A and B included in access estimates including access points, rights of way and facilities such as benches (GBS C excluded). (3) Access measures include GBS quality, size, function and weighted network distances (d) to access points. (4) Longitudinal exposure estimates created through quarterly repeat analysis. (5) GBS, green-blue spaces.

SWB, for a national population. This will be fulfilled through a number of objectives within two work packages (WPs). WP1 will use routinely collected health and demographic data (data collected for purposes other than research⁴⁸) stored in an anonymised databank to examine the risk of CMD using longitudinal changes in access to neighbourhood GBS. WP2 will link routine data in the anonymised databank with in-depth survey responses from the data linked National Survey for Wales (NSW) to evaluate a representative sample of Welsh residents' self-reported well-being and GBS use.

Objectives

WP1

1. Create a longitudinal dataset of GBS in Wales using UK Ordnance Survey (OS), local authority and remotely sensed satellite data.
2. Create longitudinal, residential GBS accessibility indices for all homes in Wales using the longitudinal GBS dataset in a Geographic Information System (GIS).
3. Create an 11-year dynamic cohort of individual-level longitudinal residential GBS accessibility indices to answer the research question(s): 'Do people with different GBS access through time have different associated risks of having a CMD?' and 'Is the association between changes in access and exposure to GBS and CMD modified by multiple sociophysical variables, migration and socioeconomic disadvantage?'

WP2

1. Create data linkages between survey and routinely collected data within the Secure Anonymised Information Linkage (SAIL) for household-level GBS accessibility indices, and individual-level health and demographic data.
2. Complement residential GBS accessibility indices by including GBS usage from the data-linked NSW to model interactions to answer research question: 'Are there associations between residential GBS access and exposure, and well-being and CMDs? Are these associations modified by GBS use and multiple sociophysical modifiers?' We will consider stratifying by CMD to check for reverse causation.

METHODS AND ANALYSIS

Design

The GBS project is a retrospective and controlled, population-wide study. We will evaluate the association between changes in access and exposure to GBS, on the risk of CMD (WP1) and SWB (WP2). This could be either at environment level (eg, a change in the GBS itself) or person level (eg, moving home, better access to transport). WP1 will use longitudinal data to examine variability in time and space of access and exposure to GBS and examine whether this could be due to planning and environmental policies. WP2 will use cross-sectional data to investigate whether visits to GBS improve SWB.

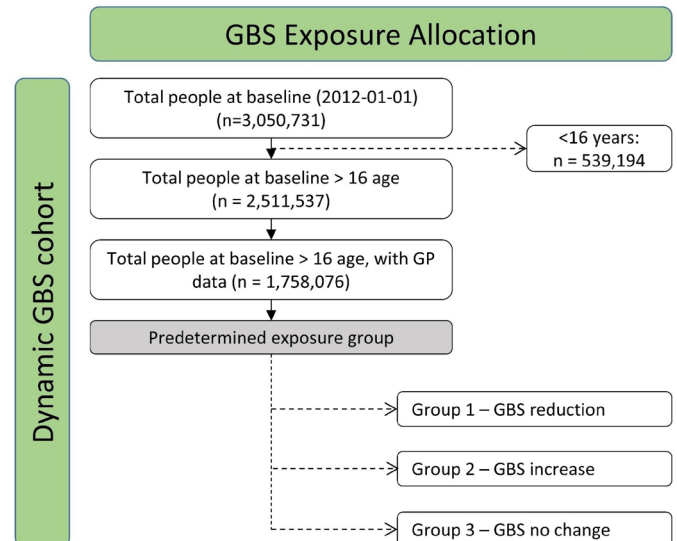


Figure 2 Flow diagram of proposed number (1.7 million adults) we currently have before exposure group allocation. GBS, green-blue spaces; GP, general practitioner.

Participants

Our study population contains people aged 16 years and older living in Wales, UK. WP1 includes the total adult population registered with a general practitioner (GP), providing GP records to SAIL. This is expected to be about 1.7 million adults in Wales (see figure 2). WP2 includes a representative sample of the adult population in Wales based on the NSW for 2 years (cross-sectional samples in 2016/2017 and 2017/2018). The NSW has an annual sample of approximately 12 000 responders and GBS visit questions are asked of 50% of that sample. This provides a total cross-sectional sample of 12 000 (over the 2 years) for this part of the study.

Research environment

We will use a secure GIS platform hosted by United Kingdom Secure e-Research Platform (UKSeRP).⁴⁹ High-resolution map data from the OS are stored in the GIS, including point data for all residences (AddressBase Premium), and road network data.^{50 51} Environment data will be collected, stored and processed to generate household - level GBS accessibility indices.

In parallel with, but separate from the GIS, is the SAIL databank.⁵²⁻⁵⁴ Data linkage completed at our Trusted Third Party, will enable a longitudinal cohort to be created before statistical analysis is conducted within SAIL. The SAIL Databank contains longitudinal health, social and education data for more than 5 million people, which includes the current population of Wales, UK (3.1 million people) at any one time. The databank includes over 15 billion anonymised records and was designed to overcome data sharing issues.

A strength of the SAIL platform is the method for anonymising all individuals and households in Wales. Data are anonymised through the split file process.^{52 53} Whereby the dataset is separated into identifiable (eg, address) and

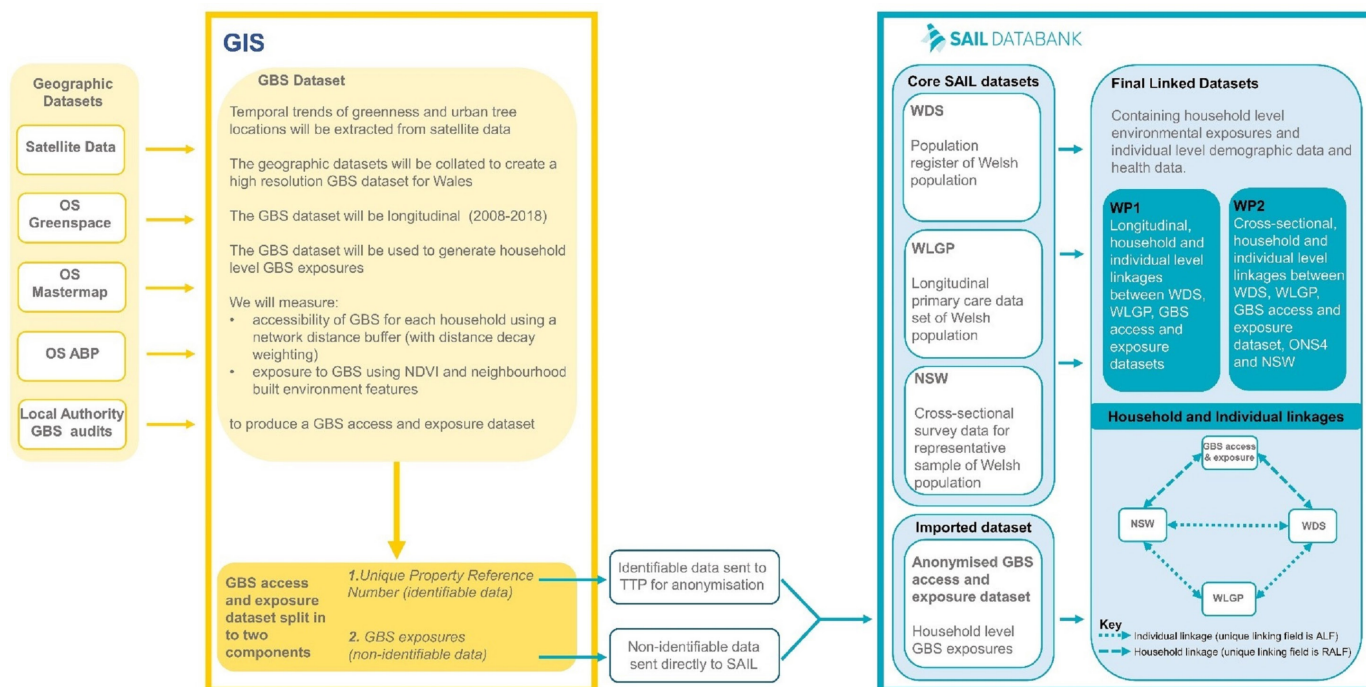


Figure 3 Diagram of proposed GIS data preparation and data linkage in SAIL databank. GBS accessibility and exposure indices created within a secure GIS platform will be linked at household level with demographics from Welsh Demographic Service (WDS) and National Survey for Wales (NSW). Common mental health conditions from GP records will be linked at individual level with the NSW and WDS. Linkage will be undertaken by our TTP (NHS Wales Informatics Service). ABP, AddressBase Premium; ALF, Anonymised Linking Field; GBS, green-blue spaces; GP, general practitioner; GIS, Geographic Information System; NDVI, Normalised Difference Vegetation Index; ONS, Office for National Statistics; OS, Ordnance Survey; RALF, Residential Anonymised Linking Field; SAIL, Secure Anonymised Information Linkage; TTP, Trusted Third Party; WDS, Welsh Demographic Service; WLGP, Welsh Longitudinal General Practice; WP1, work package.

non-identifiable (GBS access and exposure) components. The identifiable component is transported to our trusted third party (TTP), NHS Wales Informatics Service.⁵⁵ The non-identifiable component is sent securely to the SAIL Databank. The TTP anonymise and encrypt the identifiable data and each individual record is assigned a unique linking field. An Anonymised Linking Field (ALF) is assigned to individual level data and a Residential Anonymised Linking Field is assigned to a place of residence.⁵⁵ The anonymised elements of the dataset are then sent to SAIL to be loaded. These elements are then recombined with the non-identifiable (GBS access and exposure) component of the dataset, which makes them ready for linkage with other datasets in the SAIL Databank (see figure 3).

Data sources

Environment datasets

We will create a longitudinal, national dataset of GBS for Wales (2008–2018) using temporally and spatially referenced satellite imagery.⁵⁶ We will also develop a typology of GBS and collate data from multiple sources to augment the all-Wales OS data we hold in our GIS database (OS Mastermap).^{50 51 57} Using satellite imagery, we will extract temporal trends of ‘greenness’ (using measures such as the Normalised Difference Vegetation Index (NDVI), which has been used in previous studies⁵⁸) for every

household in Wales and we will also explore the integration of new GBS and land use data as they become available and are published (eg, OS Greenspace data⁵⁹) and local authority audits of GBS.

GBS accessibility and exposure dataset

There is no clear consensus in the existing literature on the most appropriate measures of GBS access and/or exposure.⁴² Using the aforementioned environment datasets, this project will collate a set of measures using criteria including longitudinal consistency, spatial resolution and evidence for associations with mental health and SWB outcomes. We will create residential-level exposure measures by modelling the ambient home environment, that is, the GBS immediately surrounding their home and what they see out of their window. This metric will include measures such as greenness (NDVI), urban trees and garden size. The access measures will represent people’s potential to visit GBS in their locality. We will adapt a previous methodology that modelled ‘change in alcohol outlet density and alcohol-related harm to population health’ (CHALICE).⁶⁰ The distance decay for this accessibility model will be updated to emulate how people engage with GBS as opposed to alcohol outlets. People will behave differently in accessing GBS, compared with how they may access alcohol outlets. For example, people may be prepared to walk or drive further to access some

green blue spaces compared with alcohol outlets.⁶¹ Our modelling will consider differential associations for key subgroups (eg, deprived populations) thereby minimising risks that recommendations could increase inequalities. We will create a number of accessibility and exposure measures, allowing planners to consider the configuration (size, function, quality) of the most beneficial mix of spaces. We will also measure, for each adult, change in GBS access and exposure for the study period. Again, we will build on the CHALICE study⁶² that measured change in alcohol outlet density by calculating change in density between the current and previous quarter and also the change between each quarter and baseline.

The longitudinal, household-level, GBS accessibility and exposure dataset will be imported into the SAIL databank and linked with health datasets that are held within the SAIL Databank. Household identifiers (Unique Property Reference Numbers, UPRNs) will be used to distinguish each household in the GIS and will have attached to each a GBS accessibility and exposure index (see figure 3). Household identifiers will be replaced with a Residential Anonymised Linking Field within SAIL.⁶³

Health datasets

Welsh Longitudinal General Practice

The Welsh Longitudinal General Practice (WLGP) dataset contains individual-level health data including Read codes⁶⁴ for all diagnoses, symptoms and treatments recorded by a GP and we will extract outcome data for each person from 2008 to 2018.

Welsh Demographic Service

The Welsh Demographic Service (WDS) dataset contains addresses for all individuals who register with a GP. Dates for each address record update are recorded, thereby providing durations of residency for multiple homes and the ability to link to local environment indices at each time point. This dataset holds demographic data including age and gender. These data will be used to create population subgroups based on age, gender and location for each period. The WDS contains historical patient provided address information linked anonymously at the individual level (the ALF) which is the primary key variable for record linkage.^{52 53}

National Survey for Wales

The NSW is a cross-sectional, representative sample of adult participants across Wales. Participants answer questions on public services such as education, transport, leisure activities, and self-reported health and well-being. They also report information on visits to GBS, such as visit frequency and activities undertaken during visits.

Figure 3 shows how GBS accessibility indices will be generated and linked with routinely collected health and survey data held within the SAIL Databank. Table 1 shows the data sources and variables that we plan to derive from the routinely collected health data, survey dataset,

Table 1 Data sources for outcome and accessibility/exposure variables to be used in statistical analysis for years 2008–2018

Dataset name	Data source	Derived variables	Coverage	Work package (WP)
GBS accessibility and exposure dataset.	Local authorities, Ordnance Survey, satellite data sources.	Typology to be developed. The dataset will include GBS such as parks, woodland, sports facilities, trees in residential areas.	Wales.	WP1 and WP2
Welsh Longitudinal General Practice dataset (core SAIL dataset).	Primary care records recorded by GP.	CMDs such as anxiety and depression (see outcome variable section for details).	70% of practices in Wales provide data to SAIL.	WP1 and WP2
Welsh Demographic Service dataset (core SAIL dataset).	NHS Wales Informatics Service.	Age, gender, week of birth, multiple move in and out of home dates.	Total population of Wales who are registered with a GP, a free to use service in the UK.	WP1 and WP2
National Survey for Wales (restricted SAIL dataset).	Welsh Government.	Subjective well-being, self-reported visits to GBS.	Cross-sectional, representative sample approximately 12 000 participants per year.	WP2

GBS, green-blue spaces; GP, general practitioner; CMDs, common mental health disorders; SAIL, Secure Anonymised Information Linkage.

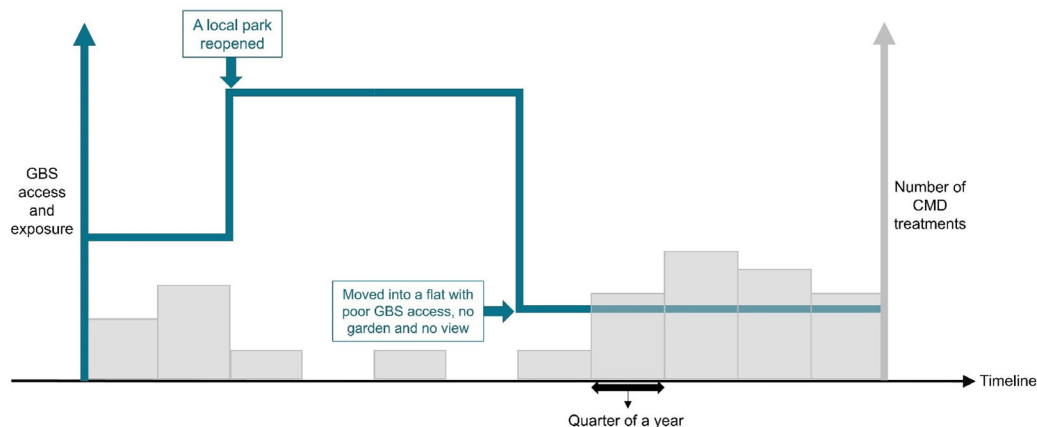


Figure 4 Conceptual model of primary exposure variable and primary outcome variable. CMD, common mental health disorder; GBS, green-blue spaces.

and population register that are available in the SAIL Databank.

STATISTICAL ANALYSIS

Outcome variables

WP1 and 2: primary outcome - CMD

Change in counts of CMD treatments for adults identified as CMD cases within the corresponding time periods for the 70% of adults in Wales for whom we have GP data records in SAIL (1.7M adults). Prevalence algorithms,^{65 66} detect cases of CMD (anxiety and depression) from routinely collected GP data. We will use the algorithm that incorporates: an historical diagnosis and currently treated, with current diagnosis whether treated or untreated, to identify CMD cases. For each adult, and for the time periods (quarters) identified as a CMD case, treatments will be counted per day and aggregated into quarterly counts. **Figure 4** shows a conceptual model of exposure variable (GBS access and exposure) and primary outcome variable (CMD treatments).

WP2: secondary outcome - SWB

SWB is measured by the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the NSW for two survey years (2016/2017 and 2017/2018) for a representative population.⁶⁷ The WEMWBS is comprehensive (incorporating elements of both SWB and psychological well-being), short enough to be used in population-level surveys, responsive to change and has been validated among community samples of adults in the UK.⁶⁸ WEMWBS scores are on continuous scale from 14 to 70.

WP1: secondary outcome - GP events

We have the total number of events recorded for each person in a WLGP dataset. We will calculate the number of GP events, converting to a binary daily activity or no activity aggregated to quarterly counts. This eliminates double counting (eg, counting the return of large quantities of test results only once). These are purposefully all events rather than those related only to CMD because this will capture people who frequently visit their GP but for

whom no CMD diagnosis has been made. This will allow us to explore social isolation and as indicator of undiagnosed CMD.

WP2: secondary outcome - Office for National Statistics SWB

We will have individual responses in NSW to the four standard SWB questions (life satisfaction, sense of worth of activities, happiness, and anxiety) as used by Office for National Statistics and the Organisation for Economic Co-operation and Development.⁶⁹ As well as being related to national representative survey data, these data have also been used in previous GBS visit research using the comparable Monitor of Engagement with the Natural Environment for England.⁴⁵

Analysis plan

We will provide descriptive statistics of GBS access and exposure and changes in GBS access and exposure by household, Lower Layer Super Output Area (LSOA), deprivation quintiles using Welsh Index of Multiple Deprivation (WIMD) and population characteristics. NSW data will have descriptive statistics of each of the survey questions we will use (number of visits, labour market status, etc).

Across both WPs, we have outcomes of two distinct types: continuous (SWB) and counts (CMD treatments and GP event days). We intend to analyse the continuous outcome using a linear model. Poisson models will be used to analyse the longitudinal count data.

Our data are hierarchical in nature: WP1: three-level data: LSOA/Individual/Time (Quarter); WP2: two-level data: LSOA/Individual. Accordingly, we will adopt a multilevel modelling approach. We will generalise the standard linear and Poisson models to handle two or three levels of variation, as appropriate. The resulting two-level and three-level models will allow us to estimate the extent to which variation at each level may be explained by confounding variables (see **table 2**). Having controlled for confounders, we will proceed to add access and exposure variables to our models, thereby allowing us to test for the significance of the access and exposure

Table 2 Analysis summary (P: primary; S: secondary)

WP	Outcomes (type)	Exposures	Confounders	Interactions of interest
1	P: CMD (count)	P: Access and exposure to GBS.	LSOA level: Quintile of deprivation (WIMD). Category of urban/rural settlement type (ONS classification).	Change in access and exposure to GBS by deprivation.
	S: GP event days (count)	S: Time of move(s).	Individual level (SAIL): gender, age, comorbidities.	Change in access and exposure to GBS by time of move(s).
2	P2: SWB (continuous)	P2: Access and exposure to GBS.	LSOA level: as per WP1.	Exact number of visits by individual level deprivation.
	P1: CMD (count)	S: Level of engagement in 150+ min of moderate or vigorous intensity activity per week.	Individual level and NSW: gender, age, comorbidities, highest educational qualification, marital status (incl. living with partner). Exact number of visits made outdoors for recreation in last 4 weeks.	Exact number of visits by residential GBS.

CMD, common mental health disorder; GBS, green-blue spaces; GP, general practitioner; NSW, National Survey for Wales; ONS, Office for National Statistics; SAIL, Secure Anonymised Information Linkage; SWB, subjective well-being; WP1, work package 1.

variables in the presence of the confounders. Inclusion of these variables in the above models will permit us to estimate their direct effects on our outcomes. Having included the main effects of both confounders and access and exposure variables, we will proceed to test for the significance of selected pairwise interactions of interest, as shown in [table 2](#).

DISCUSSION

This study will be the first of its kind to link GIS-generated GBS accessibility and exposure measures with routinely collected, longitudinal health data and cross-sectional survey data for a whole population. We plan to create a longitudinal GBS dataset for Wales. Using multisource data, we will build a dataset that records local-level changes in GBS for Wales, for 11 years. Longitudinal studies that have previously examined GBS exposure have used cross-sectional environment data to calculate GBS accessibility and exposure indices. This may be because it is a resource and time-intensive task to create a longitudinal GBS dataset, and longitudinal data are not always available. In addition, bringing together data from different sources, harmonising the data and deriving GBS indices from the data requires expertise and specialised skills. A national study using routinely collected data on a national scale is timely following recommendations from a report on a prospective quasi-experimental study.⁴⁷

Our study will work with stakeholders and policy-makers to develop a GBS typology that can be used to provide evidence that can be translated to help policy and practice. Improved evidence on the impacts of GBS on mental health is required to inform decisions relating to planning, area regeneration and environmental management.

Natural England, England's statutory body responsible for the natural environment, concluded that a knowledge of causal pathways and contributory mechanisms that link mental health and environmental exposure is required.⁷⁰ This study will produce the evidence needed to address gaps stated by the Welsh Government Environment Bill White Paper⁷¹ and inform implementation of the Well-being of Future Generations (Wales) Act 2015.⁷² We have key policy-makers signed up as stakeholders for this project and they will be involved in discussions for how we define GBS, how we measure access and exposure to GBS and in the reporting of our results so that this study can be useful in generating evidence-based policy.

Another aim of the study is to create novel GBS accessibility indices to use network routes from a variety of sources, and access points, to model how people may access GBS. The way that studies measure access to GBS is methodologically diverse and there is no general consensus on which is the best measure to use. Our measures will be informed by theories and findings in the literature. To complement this, we will ensure we involve members of the public in our research. The overarching aim of our longitudinal study is to identify causal mechanisms to determine whether a positive change in access and exposure to GBS lowers the risk of CMD.

Patient and public involvement

During the development of the study exposure measurement, we engaged with the SAIL Data-bank consumer panel, who provided us with public perspectives. . We will invite members of the public to workshops and ask them to help direct our research through a series of focused questions. The workshop group will comprise members who are experienced at considering the value of environment from the Health and Environment Public Engagement (HEPE) Group

(hosted by University of Exeter), who will join Wales-based members of the public from urban park groups, and those experienced in considering data linkage proposals.⁷³ All members will be updated with project progress.

Ethics and dissemination

Ethical arrangements

This study is based on routinely collected administrative, environment and survey data. All data will be anonymised into a secure databank, and therefore, there will be no mechanism for informing potential study participants of possible benefits and known risks. We have obtained informed consent to use the linked and anonymised NSW data within the SAIL databank. All routinely collected anonymised data held in SAIL are exempt from consent due to the anonymised nature of the databank (under section 251, National Research Ethics Committee (NREC)).

Research governance

We have applied and been granted approval by the independent Information Governance Review Panel (IGRP) for permission to conduct this study (study number 0562). The IGRP contains independent members from NREC and British Medical Association (BMA), as well as lay members, and have previously given permission for similar projects (eg, NIHR PHR CHALICE and NIHR PHR Carmarthenshire Housing). The review process has checked that the study we are is useful, not service evaluation, and will not break anonymisation standards.

Dissemination policy

We will regularly report our progress to the study steering committee (SSC). The SSC will comprise academic experts and stakeholders (NRW, Keep Wales Tidy, Sport Wales and Welsh Government). At the end of the study, we will hold a workshop to report our findings to stakeholders and the public. We will disseminate our findings to patient, policy and academic networks (eg, Health Data Research UK, Administrative Data Research and National Institute for Health Research) and we will present results to the public with easily accessible media (eg, using infographics) to maximise international engagement. We will present findings via seminars to key health professionals, including Public Health England and Public Health Wales, health service commissioners, local authorities and government planning officials to make recommendations for future policy decisions in this area and to those who have an interest in improving GBS, CMD and promoting SWB. We also plan to publish papers in internationally peer-reviewed journals to disseminate the research to the wider academic community and add to the evidence base. We will share our results at national and internationally recognised conferences and promote our findings in academic circles.

Author affiliations

¹Swansea University Medical School, Swansea University, Swansea, UK

²European Centre for Environment and Human Health, University of Exeter Medical School, Knowledge Spa, Royal Cornwall Hospital, Cornwall, UK

³Instituto de Salud Global de Barcelona.c/ Rosselló, 132, 5^a 2^a, Barcelona, Spain

⁴Research Centre in Applied Sports, Technology Exercise and Medicine, College of Engineering, Swansea University, Swansea, UK

⁵DECIPHer, Centre for Trials Research, Cardiff University, Cardiff, UK

⁶Department of Public Health and Policy, University of Liverpool, Liverpool, UK

Acknowledgements This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) databank. We would like to acknowledge all the data providers who make anonymised data available for research. We would like to acknowledge SAIL databank consumer panel for providing us with the public perspectives.

Contributors AM and JS contributed equally to writing this paper. This paper was based on a grant proposal developed by SER, RF, AA, DB, RJ, BWW, RL, RAL, MN, GS, JW and MW. This paper was written based on a draft manuscript written by RF based on the grant proposal. Specifically, the Introduction and discussion were written by AM. JS wrote the analysis plan and developed the conceptual model. The methods section was written by AM and JS. All the coauthors made substantial contributions to the writing of the methods section and editing of the whole paper. All authors then read the final version and approved it for submission and publication.

Funding This project was funded by the NIHR PHR (project number 16/07/07). This work was supported by Health Data Research UK (NIWA1), which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, National Institute for Health Research (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. RF is supported by the National Centre for Population Health and Wellbeing Research (NCPHWR).

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. World Health Organization. *Depression and other common mental disorders: global health estimates*: World Heal Organ, 2017:1–24.
2. Davies S. *Annual Report of the Chief Medical Officer 2013. Public Mental Health priorities: investing in the evidence*. London: Department of Health, 2013.
3. Centre For Mental Health. Economic and social costs of mental health problems. <https://www.centreformentalhealth.org.uk/economic-and-social-costs> (Accessed 13 Jun 2018).
4. Diener E, Chan MY. Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Appl Psychol* 2011;3:1–43.
5. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *Lancet* 2015;385:640–8.
6. WHO. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med* 1995;41:1403–9.
7. Hartig T, Mitchell R, de Vries S, et al. Nature and health. *Annu Rev Public Health* 2014;35:207–28.
8. Marmot M. Fair society, healthy lives.. 2010.
9. Markevych I, Schoierer J, Hartig T, et al. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environ Res* 2017;158:301–17.
10. Dadvand P, de Nazelle A, Figueras F, et al. Green space, health inequality and pregnancy. *Environ Int* 2012;40:110–5.

11. Bowler DE, Buyung-Ali LM, Knight TM, *et al.* A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health* 2010;10:456.
12. James P, Banay RF, Hart JE, *et al.* A Review of the Health Benefits of Greenness. *Curr Epidemiol Rep* 2015;2:131–42.
13. Thompson Coon J, Boddy K, Stein K, *et al.* Does participating in physical activity in outdoor natural environments have a greater effect on physical and mental wellbeing than physical activity indoors? A systematic review. *Environ Sci Technol* 2011;45:1761–72.
14. Coombes E, Jones AP, Hillsdon M. The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Soc Sci Med* 2010;70:816–22.
15. White MP, Alcock I, Wheeler BW, *et al.* Would You Be Happier Living in a Greener Urban Area? A Fixed-Effects Analysis of Panel Data. *Psychol Sci* 2013;0956797612464659.
16. White MP, Alcock I, Wheeler BW, *et al.* Coastal proximity, health and well-being: results from a longitudinal panel survey. *Health Place* 2013;23:97–103.
17. Hystad P, Davies HW, Frank L, *et al.* Residential greenness and birth outcomes: evaluating the influence of spatially correlated built-environment factors. *Environ Health Perspect* 2014;122:1095–102.
18. Davdand P, Sunyer J, Basagaña X, *et al.* Surrounding greenness and pregnancy outcomes in four Spanish birth cohorts. *Environ Health Perspect* 2012;120:1481–7.
19. Kingsley SL, Eliot MN, Whitseal EA, *et al.* Maternal residential proximity to major roadways, birth weight, and placental DNA methylation. *Environ Int* 2016;93:43–9.
20. de Vries S, van Dillen SM, Groenewegen PP, *et al.* Streetscape greenery and health: stress, social cohesion and physical activity as mediators. *Soc Sci Med* 2013;94:26–33.
21. van den Berg MM, van Poppel M, van Kamp I, *et al.* Do physical activity, social cohesion, and loneliness mediate the association between time spent visiting green space and mental health? *Environ Behav* 2017;001391651773856.
22. Maas J, van Dillen SME, Verheij RA, *et al.* Social contacts as a possible mechanism behind the relation between green space and health. *Health Place* 2009;15:586–95.
23. Richardson E, Pearce J, Mitchell R, *et al.* The association between green space and cause-specific mortality in urban New Zealand: an ecological analysis of green space utility. *BMC Public Health* 2010;10:240.
24. Mitchell R, Astell-Burt T, Richardson EA. A comparison of green space indicators for epidemiological research. *J Epidemiol Community Health* 2011;65:853–8.
25. Mitchell R, Popham F. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet* 2008;372:1655–60.
26. Sugiyama T, Villanueva K, Knuiaman M, *et al.* Can neighborhood green space mitigate health inequalities? A study of socio-economic status and mental health. *Health Place* 2016;38:16–21.
27. Amoly E, Davdand P, Fornis J, *et al.* Green and blue spaces and behavioral development in Barcelona schoolchildren: the BREATHE project. *Environ Health Perspect* 2014;122:1351–8.
28. Astell-Burt T, Mitchell R, Hartig T. The association between green space and mental health varies across the life course. A longitudinal study. *J Epidemiol Community Health* 2014;68:578–83.
29. Mitchell R, Popham F. Greenspace, urbanity and health: relationships in England. *J Epidemiol Community Health* 2007;61:681–3.
30. Reklaitiene R, Grazuleviciene R, Dedele A, *et al.* The relationship of green space, depressive symptoms and perceived general health in urban population. *Scand J Public Health* 2014;42:669–76.
31. de Vries S, Verheij RA, Groenewegen PP, *et al.* Natural environments—healthy environments? an exploratory analysis of the relationship between greenspace and health. *Environ Plan A* 2003;35:1717–31.
32. van den Berg M, Wendel-Vos W, van Poppel M, *et al.* Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. *Urban For Urban Green* 2015;14:806–16.
33. McEachan RRC, Yang TC, Roberts H, *et al.* Availability, use of, and satisfaction with green space, and children's mental wellbeing at age 4 years in a multicultural, deprived, urban area: results from the born in Bradford cohort study. *Lancet Planet Heal* 2018;2:e244–54.
34. Dai D. Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? *Landsc Urban Plan* 2011;102:234–44.
35. Coolen H, Meesters J. Private and public green spaces: meaningful but different settings. *J Hous Built Environ* 2012;27:49–67.
36. Taylor L, Hochuli DF. Defining greenspace: Multiple uses across multiple disciplines. *Landsc Urban Plan* 2017;158:25–38.
37. Wheeler BW, Lovell R, Higgins SL, *et al.* Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. *Int J Health Geogr* 2015;14:17.
38. Alcock I, White MP, Lovell R, *et al.* What accounts for 'England's green and pleasant land'? A panel data analysis of mental health and land cover types in rural England. *Landsc Urban Plan* 2015;142:38–46.
39. MacKerron G, Mourato S. Happiness is greater in natural environments. *Glob Environ Chang* 2013;23:992–1000.
40. van Dillen SME, de Vries S, Groenewegen PP, *et al.* Greenspace in urban neighbourhoods and residents' health: adding quality to quantity. *J Epidemiol Community Health* 2012;66:1–5.
41. Annerstedt M, Ostergren PO, Björk J, *et al.* Green qualities in the neighbourhood and mental health - results from a longitudinal cohort study in Southern Sweden. *BMC Public Health* 2012;12:337.
42. Ekkel ED, de Vries S. Nearby green space and human health: evaluating accessibility metrics. *Landsc Urban Plan* 2017;157:214–20.
43. Grelhier J, White MP, Albin M, *et al.* BlueHealth: a study programme protocol for mapping and quantifying the potential benefits to public health and well-being from Europe's blue spaces. *BMJ Open* 2017;7:1–10.
44. Gascon M, Triguero-Mas M, Martínez D, *et al.* Mental health benefits of long-term exposure to residential green and blue spaces: a systematic review. *Int J Environ Res Public Health* 2015;12:4354–79.
45. White MP, Pahl S, Wheeler BW, *et al.* Natural environments and subjective wellbeing: different types of exposure are associated with different aspects of wellbeing. *Health Place* 2017;45:77–84.
46. White MP, Pahl S, Ashbullby K, *et al.* Feelings of restoration from recent nature visits. *J Environ Psychol* 2013;35:40–51.
47. Ward Thompson C, Silveirinha de Oliveira E, Tilley S, *et al.* Health impacts of environmental and social interventions designed to increase deprived communities' access to urban woodlands: a mixed-methods study. *Public Health Res* 2019;7:1–172.
48. Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. *CMAJ* 2017;189:E1054–5.
49. Health Informatics Group SUMS. *UK Secure eResearch Platform UKSeRP Data Sharing : Why Should We Share Data?* Swansea: Swansea University Medical School, 2017.
50. Ordnance Survey. AddressBase Premium-Business and government. <http://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-premium.html> (Accessed 10 Mar 2014).
51. Ordnance Survey. OS(R) MasterMap(R) Integrated Transportation NetworkTM Layer technical specification. Crown copyright. 2011. <http://www.ordnancesurvey.co.uk/oswebsite/support/products/os-mastermap/itm-layer-technical-specification/index.html%5Cnhttp://www.ordnancesurvey.co.uk/oswebsite/docs/user-guides/points-of-interest-user-guide.pdf%5Cnhttp://www.ordnancesurvey.co.uk/oswebsi>
52. Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;9:3.
53. Ford DV, Jones KH, Verplancke JP, *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9:157.
54. Rodgers SE, Lyons RA, Silva R, *et al.* Residential Anonymous Linking Fields (RALFs): a novel information infrastructure to study the interaction between the environment and individuals health. *J Public Health* 2009;31:582–8.
55. SAIL Databank. The Secure Anonymised Information Linkage Databank. <https://saildatabank.com/saildata/data-privacy-security/#protecting-identities> (Accessed 3 Oct 2017).
56. Gan M, Deng J, Zheng X, *et al.* Monitoring urban greenness dynamics using multiple endmember spectral mixture analysis. *PLoS One* 2014;9:e112202.
57. Ordnance Survey. OS MasterMap Topography Layer. 2017. <https://www.ordnancesurvey.co.uk/business-and-government/products/topography-layer.html> (Accessed 15 Nov 2017).
58. Beyer KM, Kaltenbach A, Szabo A, *et al.* Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. *Int J Environ Res Public Health* 2014;11:3453–72.
59. *OS Greenspace | OS GetOutside.*
60. Fry R, Orford S, Rodgers S. A best practice framework to measure spatial variation in alcohol availability. *Environ Plan B Plan Des* 2018;1–19.
61. Fone D, Morgan J, Fry R, *et al.* Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): a comprehensive record-linked database study in Wales. *Public Health Res* 2016;4:1–184.

62. Fone D, Dunstan F, Lee S, *et al.* Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): a comprehensive record-linked database study in Wales. 2015. <http://www.nets.nihr.ac.uk/projects/phr/09300702> (Accessed 24 Jun 2015).
63. Rodgers SE, Lyons RA, Dsilva R, *et al.* Residential Anonymous Linking Fields (RALFs): a novel information infrastructure to study the interaction between the environment and individuals' health. *J Public Health* 2009;10:1-7.
64. NHS Digital. Read Codes. <https://digital.nhs.uk/article/1104/Read-Codes> (Accessed 19 Dec 2017).
65. John A, McGregor J, Fone D, *et al.* Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;16:35.
66. Cornish RP, John A, Boyd A, *et al.* Defining adolescent common mental disorders using electronic primary care data: a comparison with outcomes measured using the CIS-R. *BMJ Open* 2016;6:e013167.
67. Aumeyr M, Brown Z, Doherty R, *et al.* National Survey for Wales 2016-17 Technical Report (Updated) National Survey for Wales 2016-17 Technical Report. 2017. http://doc.ukdataservice.ac.uk/doc/8301/mrdoc/pdf/8301_171018-national-survey-wales-2016-17-technical-report-en.pdf (Accessed 19 Jun 2018).
68. Tennant R, Hiller L, Fishwick R, *et al.* The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes* 2007;5:63.
69. Office for National Statistics. Personal Well-being user guidance. <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide> (Accessed 30 Aug 2018).
70. Natural England. Links between natural environments and mental health: evidence briefing - EIN018. 2016. publications.naturalengland.org.uk/file/468126390707814
71. Welsh Government. White Paper Towards the Sustainable Management of Wales' Natural Resources, Consultation on proposals for an Environmental Bill, WG19631. 2013.
72. Welsh Government. Well-being of Future Generations (Wales) Act 2015. <http://gov.wales/topics/people-and-communities/people/future-generations-act/?lang=en> (Accessed 7 Nov 2017).
73. SAIL Databank. Consumer Panel. <https://saildatabank.com/about-us/public-engagement/> (Accessed 12 Jul 2018).