Linking transport, health and sustainability: Better data sets for better policy-making

CrossMark

Clemence Cavoli [a,*], Nicola Christie [a], Jennifer Mindell [b,1], Helena Titheridge [a]

[a] Civil, Environmental & Geomatic Engineering, University College London, Gower Street, WC1E 6BT London, UK
[b] Epidemiology & Public Health, University College London, HSSRG, Department Epidemiology & Public Health, 1–19 Torrington Place, London WC1E 6BT, UK

ABSTRACT

The impact transport has on our physical and mental health and on the environment is increasingly recognised by academics, practitioners and decision-makers. To inform policy-making and research, it is crucial to have access to sufficient comprehensive datasets linking these topics. Large scale surveys rarely combine questions on transport, health and sustainability, limiting their usefulness in research and policy-making. This project set out to identify the gaps in administrative and survey datasets to link health outcomes to travel behaviour and the quality of the environment; investigate the impact this has on research and policy-making; assess how these gaps might be addressed; and identify what the needs are for joined datasets. To achieve this, the project interviewed key decision-makers and stakeholders across England, including civil servants within national or local government and third sector organisations. The results highlight gaps within national datasets; the insufficient links between health, transport, and sustainability datasets; and the need for further joined data in various fields, in particular active travel, health and wellbeing. Participants suggested several solutions, including better harmonisation of surveys and data fusion.

1. Background

1.1. Is there a need for joined up research?

Measuring multidirectional relationships between health outcomes and the way we travel is becoming increasingly important. For instance, cycling as a way of commuting is increasing in many cities (Transport for London, 2011), but how is this affecting injury rates? Does it improve physical and/or mental health? Are the benefits of cycling outweighed by issues such as air pollution or injury?

Recent research tackles some of these questions and highlights the importance of cross-disciplinary analysis. For example, mounting evidence focuses on the link between active travel, in particular cycling and walking, and reduced health issues such as obesity (Donaldson, 2004; Mindell et al., 2011a; Woodcock et al., 2011). Studies have shown that the health benefits of cycling outweigh the risks (De Hartog et al., 2010; Hillman, 1993). The economic impact of active travel has also been studied. Some researchers have estimated that the National Health Service (NHS) in England and Wales could save up to £17 billion within 20 years if active travel were to increase (Jarrett et al., 2012). Links have also been established across the fields of environmental sustainability,[2] health and transport. Correlations between low carbon and active travel policies have been identified (Mindell et al., 2011b). Research has also highlighted the impact that environmental factors or the 'built environment' (e.g. neighbourhood, school, street) can have on people's physical and mental health (Booth et al., 2005; Law et al., 2007; Srinivasan et al., 2003). For instance, obesity or diabetes are associated with the level of deprivation or the walkability of an area (Booth et al., 2005; Law et al., 2007) and disadvantaged areas also tend to have high rates of collisions (Christie et al., 2007; Edwards, 2006; Graham et al., 2005; Watkins, 2012).

---

* Corresponding author. Tel.: +44 2076791590/7526169742.
E-mail address: clemence.cavoli.09@ucl.ac.uk (C. Cavoli).

The quality of data collected in national surveys in England has also been questioned (Mindell et al., 2012; Ward et al., 2002; Woodcock et al., 2013). Problems with cycling and walking statistics include over-counting of serious travel-related cycling injuries in hospital data, and not counting pedestrian falls as transport injuries (Mindell et al., 2012; Ward et al., 2002; Woodcock et al., 2013). Many researchers highlight that data about cycling and walking need to be better collected (Mindell et al., 2012) or better analysed (Bhatia and Wier, 2011; Mindell et al., 2012).

In the field of health and active travel, several gaps have also been highlighted. The former National Obesity Observatory identified key issues related to physical activity data collection (National Obesity Observatory, NHS, 2009). That report considered many surveys, including the Health Survey for England, Active People Survey, and the National Travel Survey and identified that there is insufficient data about the impact cycling and walking has on health (National Obesity Observatory, NHS, 2009; Woodcock et al., 2011). In the field of obesity, studies highlighted the need for more research on obesity and the built environment (Booth et al., 2005; Dugdill et al., 2009).

Research suggests that there are many gaps and issues in datasets which need to be addressed, plus a need to join and analyse datasets in a more systematic and strategic way. However, gaps and issues related to health, environmental sustainability and transport across UK datasets have not been comprehensively identified and recommendations to fill these gaps not sufficiently addressed.

1.2. Aims of the study

This project set out to identify the gaps in datasets to link transport, health, and environmental sustainability; investigate the impact these have on research and policy-making; determine how these gaps might be addressed; and identify what the needs are for joined datasets.

2. Methods: key stakeholders consultation

Between October 2012 and April 2013, key stakeholders in the fields of transport, health and sustainability in England were interviewed. In total 27 people were interviewed: 22 formally, using semi-structured interviews (See Annex 1) and five informally. Prospective participants were selected purposively with a view to representing a variety of stakeholders who use national datasets in these fields. Among those interviewed, 11 individuals specialised in transport (including safety), eight in health, seven in environmental sustainability, and two were from data providers. All the participants' fields of expertise and interests crossed into at least one other of the three areas of transport, health, and sustainability (including social and economic sustainability). In terms of transport, the focus of enquiry was personal travel and mobility (defined as the ability to move between places); the choice of participants from the transport field reflected this.

Participants from the following roles/types of institutions were represented:

- 10 researchers in academia, institutes/schools and centres
- 1 consultant
- 5 local government policy makers
- 4 national policy makers
- 4 third sector organisations
- 1 parliamentary group
- 2 data providers

The semi-structured topic guide included three parts (see Annex 1). The first focused on the data sets' advantages and disadvantages, and general questions about aims and objectives, outcomes and impacts of UK national policies related to data sets. The second part was dedicated to establishing the questions, themes or topics that participants would like to see addressed in national surveys. The last part of the questionnaire collected participants' views about different solutions and ideas on whether there should be a platform for centralising surveys or information about surveys and the potential for data fusion. Finally, participants were asked to describe what their ideal survey would be and whether this survey should focus on the local or national level. Fig. 1.

The following questions were addressed in the topic guide:

- What datasets do participants use and what do they think are the weaknesses of each of those datasets?
- What are the strengths and limitations of the data sets stakeholders use?
    - What impact do these limitations have on policy-making, evaluation and research?
- To what extent is there a need for a survey that explores the interrelationship between mobility, health, and sustainability?
    - What gaps in knowledge would such a data set help address?
- What do stakeholders think about data fusion or the establishment of a platform which would centralise all surveys?
- What would the ideal survey be like?

Participants were interviewed with informed consent. Interviews were audio taped and transcribed. Informal interviews were first conducted to gain a general understanding of the situation. This helped shape the semi-structured questionnaire. Data was coded and analysed for thematic content. Key themes are exemplified in the paper using verbatim anonymised quotations from participants. Key gaps in national data sets were identified through the thematic analysis of the interviews.

To conduct the analysis of the interviews a framework matrix was developed based on the key research questions addressed by the topic guide and then populated with responses from participants. The four themes of the matrix were:

1. comparing assets and drawbacks of main surveys;
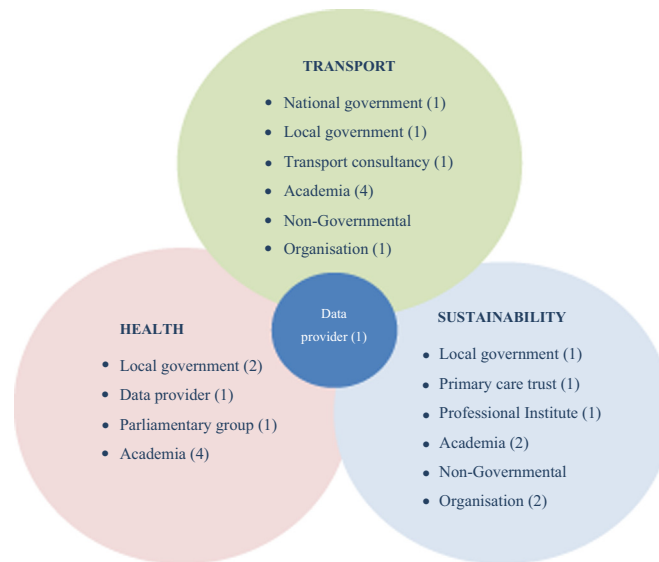2. identifying national survey gaps;

Fig. 1. List of participants.

3. investigating what questions participants would like to add to surveys; and
4. providing solutions and ideas

Each matrix was divided into sub-themes and then systematically cross-analysed to identify key themes (Miles et al., 2014). These themes were also reviewed by another researcher to check for consistency.

3. Existing data sets: limitations and impact on policy making

3.1. Gaps in national datasets

The most frequently used datasets were STATS19 (police road traffic collision dataset), Road Casualties (Department for Transport routine statistics based mainly, but not exclusively, on STATS19 data), and Hospital Episode Statistics (HES, National Health Service hospital healthcare data) - all administrative data – and the National Travel Survey (NTS), Active People Survey, and the Health Survey for England – all random sample population surveys. A brief description of each survey is shown in Table 1.

Several gaps were highlighted by the participants, as illustrated in Table 2:

The majority of the participants mentioned that one of the main limitations of current national data sets is the *limited sample size*, which does not provide data on a small geographical scale. As one of the stakeholders stated: too often '*assumptions that are made from the national datasets cannot be applied to the local population*' (participant 8, Health). London policy makers commented that in the case of London, it is such a '*unique case that national survey results are not necessarily representative*' (participant 1, Health and Transport). More generally, data sets are not '*fine grained enough*' to represent certain minorities, or look at certain '*specific issues*' such as unusual injuries (participant 11, Transport and Environment; participant 13, Health). Three data sources in particular were pointed out: the National Travel Survey (NTS), Active People Survey and reported road casualties.

In addition, participants mentioned that there is a *lack of longitudinal data sets* amongst national surveys. As one participant said there is a lack of '*Consistency and continuity*" (participant 17, Transport). Participants insisted that it is important and useful to have longitudinal data, collecting information from the same individuals over time and made suggestions such as establishing a '*baseline and measure every five years*' (participant 13, Health).

Data harmonisation describes the process of reconciling different measurements of the same parameter (e.g. age, ethnicity, travel time etc.) between different types of survey to make them mutually compatible. Participants commented on the *lack of harmonisation between surveys* and the different definitions and categories which are a barrier to harmonisation such as time inconsistencies and definition of basic characteristics such as age group, the definition of 'urban areas', as well as level of education. Between surveys there were inconsistencies over the definition of physical activities, with the Department of Health measuring 30 min of activity[3] whereas the Active People Surveys talk about 20 min. One of the key gaps identified by participants is that active travel is not sufficiently addressed through national surveys. Even though current data sets provide, to some extent, a big picture of the situation, more data sets are needed to fully understand and measure active travel and its impact on health. This is especially true for walking and cycling as well as active commuting under 30 min. Participants noted that current national data sets do not differentiate between cycling and walking as leisure, a way to commute, or other travel. As stated by a participant: '*Getting a holistic picture of how much walking and cycling contributes to people's overall activity and for what purpose is quite difficult.*'(participant 10, Health).

A key missing element related to walking is *insufficient data on short walking journeys* between different modes of transport. Walking as part of '*multi part journeys*' is not adequately recorded and as a result the '*value of walking is under-recognised*' (participant 13, Health). According to many participants, cycling is equally under-represented in national surveys. They felt that it was not sufficiently researched

---

[3] The recommendation has recently changed and is now 150 min of moderately vigorous activity per week.

Table 1
National surveys and routine datasets relating to health and transport in the UK.

Road casualty data 'STATS 19'	Routinely collected data on road collisions reported to the police where someone has been injured (STATS19 system). Includes information on mode of travel, age and gender of all drivers an all injured persons, circumstances of collisions, types of vehicle involved and severity of injury. Published annually.
Hospital Episode Statistics	Routinely collected data on all admissions, outpatient appointments and A&E[a] attendances at NHS hospitals in England. It is a record-based system that covers all NHS trusts in England. The database processes over 125 million records annually. Data includes diagnoses and procedures (operations), and patient demographics. Road traffic casualties can be identified through the International Classification of Disease external cause codes.
National Travel Survey	Data on personal travel patterns in Great Britain. Based on an annual household survey of approximately 20,000 people from 8000 households. Data is based on interviews and seven day travel diaries and covers how, why, when and where people travel as well as factors which affect personal travel such as car availability, driving licence holding and access to key services. Published annually.
Active People Survey	Data on the number of people taking part in sport across the nation and in local communities. Telephone based interviews among people age 14 and over throughout the year. The key measure is the '1 x 30' indicator, the percentage of the adult population participating in sport, at moderate intensity, for at least 30 minutes on at least four days out of the last four weeks. Cycling is included if done at least once a week at moderate intensity for 30 minutes. Published annually.
Health Survey for England	Data on health (e.g. Body Mass Index, longstanding illness) and health-related behaviours in adults and children living in private households in England. Published annually.

[a] A&E: Attendance at hospital Accident and Emergency department (Emergency Room).

Table 2
Gaps in National Data Sets.

Gaps in National Data Sets	
Structural gaps	**Conceptual gaps**
Insufficient sample	Active travel needs to be better addressed
Longitudinal surveys missing	The four areas are not sufficiently linked
Lack of harmonisation between surveys (different definitions)	Health and Transport lack connection

and not well recorded (see below): *'Statistics about cycling are often confusing,'* (participant 10, Health). The National Travel Survey, in particular, includes journeys only on routes along which motor vehicles can travel, and therefore undercounts walking and cycling trips and distances, given the marked rise in such journeys on traffic-free routes in the past decade (Sustrans, 2013).

Many participants highlighted the lack of *data sets which link health and transport*, the need for *'public health policies to get information from transport and* vice versa' (participant 15, Transport). Some participants stated that highlighting the relationship between travel and health is a key way to change people's behaviour and that the link between transport and obesity needs to be better highlighted. It was also mentioned that changes on the ground, such as infrastructure, are not sufficiently linked *'with actual health outcomes'* (participant 11, Transport and Environment). Other participants talked about the need to link the fields of transport, health and sustainability. As stated by one:

'We need to understand the bigger picture and determine what has the biggest impact on health. Is it air quality? Or inactivity? So if for example it was the case that it was a city where air quality was causing much greater health impacts than road traffic collisions or physical inactivity then the policy makers should be focusing their efforts on air quality. But if it's a city where actually it's physical inactivity that's the biggest impact on health, then that's where the efforts should go.'(participant 1, Health and Transport).

Several participants indicated that there are *flaws in injury data sets*. The most frequent issues are linked to reported cycling casualty rates. Datasets such as STATS19 indicate that *'the number of casualties increases but it is linked to the increase in the number of cyclists'* (participant 10, Health). Reacting to a newspaper article about the increase in cycling collisions during the Olympics one participant said: *'You've got to put that in the context of exposure stats. […] It might actually have become safer to cycle over the same period because so many people were out on their bikes.'* (participant 3, Transport and Environment).

3.2. Impact on policy making and research

The majority of participants felt that the limitations of existing datasets influence and limit research to a large extent. One participant stated that: *'We should be able to tell a compelling story about the impact of transport on health but at the moment it is a struggle for us to find basic data'* (participant 1, Health and Transport). Most participants felt that policy objectives are not sufficiently addressed by surveys, although some interviewees were unsure. Some areas were quoted as missing the data to enable the evaluation of national policy targets, particularly health, transport and carbon emissions. In the field of health and transport, there is a gap between national policies which *'are giving priorities to obesity and physical activity'* and the data sets available (participant 6, Transport). There is also a need for more data sets about cycling, since many policies have launched schemes for increased investments.

Most participants mentioned that *policy outcomes are not sufficiently measured by national data sets*. In general *'there is a lack of evaluation'* and this is especially true in certain fields such as walking, cycling, carbon emissions, air quality, behavioural changes, which are *'hardly measured',* and more generally long term impacts (participant 2, Transport and Health; participant 13, Health). Participants pointed out that evaluation is crucial, that *'Policies should be prospectively evaluated'* and that *'pre- and post-evaluation of interventions'* are very useful (participant 2, Transport and Health; participant 6, Transport). As one participant said: *'we need to know what is the value for*

money of schemes. We need to understand the value on the return of our investment'. Identifying whether investments in policies or projects benefit society is key (participant 6, Transport).

4. New data sets: users' needs

Many participants expressed their interest in linking transport, health, and sustainability; *'It is useful that those relationships are explored'* (participant 7, Health). It is important to establish *'how changes in one area, for instance health, can have a spill over other areas, such as environment.'* (participant 22, Sustainability). Participants were asked which areas or subjects need to be better addressed by existing surveys. Four key themes emerged from the content analysis, as illustrated by Table 3 below.

4.1. Physical activity, active travel and health

Participants expressed an interest in including more general questions about health. For instance, a question such as *'How is your health in general?'* could be included in many non-health surveys (participant 3, Transport and Environment). Most participants also insisted that there is a necessity to obtain more details about *physical activity* and *active travel,* especially in relation to **obesity**. There is a need to better link certain areas such as *'Physical activity and dietary combination'* or the impact active travel has on health (participant 8, Health). Issues such as the *'Impact of walking and cycling on overall activity level and links between active travel and obesity'* were raised (participant 11, Transport).

Participants showed much interest in obtaining data about overall physical activity during the entire day and in understanding people's daily routine. There is a perceived need to have a *'Physical activity diary'* for people to record their physical activities, from leisure to commuting, including short commuting trips (participant 21, Transport). Participants felt that this would help identify: *'What proportion of it* (physical activity) *is coming from work, what proportion of it is coming from leisure activities, what type of leisure activity…'* (participant 1, Health and Transport). Fears related to cycling were also mentioned. Issues such as whether *'unpleasant experiences cyclists suffer whilst they cycle prevents them from cycling? (Such as getting abuse from drivers or threats)'* were raised (participant 2, Transport and Health). Stakeholders would like to be able to *'pick up changes in views about fear of traffic or other disincentives to cycling'* (participant 13, Health).

Other questions such as 'Why do people cycle?' or 'Who cycles?' were suggested (participant 10, Health). One participant noticed: 'We have seen the cyclists increased by 20% [in London] in terms of counts on the road; but who are those people cycling?' (participant 4, Transport). Behavioural changes related to cycling were also mentioned or questions such as 'Do people's walking and cycling patterns change across the life course? What contribution is it making to their health?' Suggestions were made such as having 'cohort studies of cyclists' (participant 10, Health; participant 2, Transport and Health).

Travel to work and especially to school was mentioned many times as being a key theme which needs to be better addressed in surveys. Many participants voiced their disappointment at the fact that the Department of Education stopped collecting data about children travelling to school at a local level. It was a *'really invaluable resource'* (participant 10, Health). One participant highlighted the fact that independent mobility for children *'is the sign of a healthy society'* and also that *'habits need to be taken at a young age.'* (participant 12, Health and Transport).

A key theme raised by many participants was *measuring well-being,* especially 'psychological well-being'. Well-being as described by participants referred to psychological or mental well-being, such as levels of depression or effect on mood, as well as quality of life and happiness. Participants felt that there is a need to address and measure well-being in surveys. In relation to *well-being*, one participant stated: *'One of the social aims is not just to have a population which is sort of physically functioning in health terms but also have a sense of purpose or a sense of happiness*[…] *then logically you need metrics to assess how well you are doing on that'* (participant 3, Transport and Environment).

In addition, participants highlighted that *behavioural change* is a key issue and should be better measured. There is a focus on collecting hard data but *'that failed to take account of the nature of changes that are needed along the way when you have got long term shifts in population level behaviours.'* (participant 13, Health). Questions such as *'In the long term, who has changed behaviour, why and how?'* need to be raised (participants 4, Transport). In the field of transport it would be useful to further tackle reasons to travel as well as behavioural issues related to cycling.

Participants showed much interest in *perceived safety,* especially in relation to cycling and walking. Areas such as *'risk perception'* and *'attitudes to road safety'* need to be addressed (participant 20, Health). One participant redefined the phrase road safety: *'Casualty reduction is not road safety. Road safety is more than that, it is freedom from fear or harm of injury on or around the road environment.'* (participant 12, Health and Transport).

Table 3
Key themes which need to be better addressed.

Physical activity, active travel and health	Well-being, behaviours and attitudes	Linking mobility, health and sustainability	Background information, identifiers and underrepresented groups
General health levels	Behavioural questions, attitude and perception	Carbon emissions, air quality and energy	Basic background characteristics (age, gender)
More details about physical activity and active travel	Perceived safety	Pleasantness of the environment where people commute	More data about deprived groups and ethnicity
Cycling	Life changing event (Transition, moving home…)	Transport and the economy	People's location
Travel to school		Social and economic sustainability	Using data from new technology

4.2. Environmental sustainability

When asked which questions would you liked to see addressed in national surveys, the majority of the participants mentioned environmental sustainability. Three key issues were specifically raised: air quality, energy, and the pleasantness of the environment.

4.2.1. Air quality
According to some, the impact pollutants emissions have on health is not sufficiently tackled in national surveys. Although participants did not mention how to address this issue through national surveys, they detected a specific need for further data sets related to '*asthma and cancer in relation to particulates*' (participant 6, Transport) or the need for more solid evidence '*about the efficiency of 20 miles an hour on fuel efficiency and pollution*' (participant 12, Health and Transport).

4.2.2. Energy
Participants highlighted the need to obtaining data sets about energy savings and understanding the value of home retrofitting. In addition, one participant commented on the need to join data about electricity and gas usage with vehicle consumption. This would help to answer questions such as '*are people who drive the most also the people who have the highest electricity and gas bills?*' (participant 3, Transport and Environment). Some participants also addressed the link between health, physical activity (especially obesity) and carbon consumption (active travel versus driving a car). One concluded by saying: '*Often what is going to help people's health is also going to help the planet*' (participant 20, Health).

The '*pleasantness of the environment where people commute*' was mentioned as being a key issue (participant 3, Transport and Environment). It is also related to perception of safety and well-being, as stated by one participant, commuters '*don't necessarily want to be walking on a busy road but you might consider walking* (to go to work) *down a more pedestrianised street*' (participant 1, Health and Transport). Participants suggested addressing questions such as '*Does a change in people's environment improve their life*?' (participant 22, Sustainability).

4.3. Under-represented groups and background information

Many participants would like surveys to focus more on *age*, especially older people and young people. One participant suggested that surveys should look at '*How we can keep people active and happy in their old age (especially since people are living much longer)*' (participant 20, Health). Linking ageing and mobility was also raised.

Many participants talked about the importance of including *deprived groups* in national survey data sets. These disadvantaged or very disadvantaged groups included the disabled, those with limited mobility or accessibility issues, those on lower incomes or education levels and the socially excluded. Participants reported that it would be useful to identify how disability affects people's daily activity (participant 20, Health). It is also very important to understand the '*Health implication of isolation and marginalisation'* (participant 8, Health).

Ethnicity was also an important factor mentioned by participants. Deprivation and ethnicity are often tied together. According to some participants, ethnic, social and religious segmentation should be given more importance in national surveys; although being able to analyse data by specific ethnic groups would require a very large survey.

It was often mentioned that obtaining '*geographic positioning*' would be extremely useful (participant 22, Sustainability). Participants argued that having access to people's postcode and the location of, for example, trip destinations would provide a wealth of information such as 'where people want to go' or 'tracing people's movements' (participants 4, Transport). It is especially important in the case of STATS19 road casualty data, where the location of the collision is specified. However, participants were aware that obtaining such data '*might not be realistic*' due to anonymity issues (participants 4, Transport).

5. The ideal survey: solutions and obstacles

5.1. Data fusion: ideas and obstacles

Almost all participants mentioned that data fusion offers solutions to many problems and gaps in national data sets. Data fusion describes the process of merging existing databases into a single participant level database to provide a deeper understanding of a specific issue. It involves matching data, i.e. pairing participants on the basis of common characteristics, usually key demographics (age, gender, area, deprivation/social class, ethnicity, education). This integration of data is based on the principle that these common characteristics can reliably predict behaviour. The outcome of data fusion, like any modelling process, needs to be validated by looking at whether its predictions are accurate. In addition, careful checks need to be made to take into account differences in the measurement of matching criteria.

Expressions such as 'harmonising', 'synchronising', 'combining' or 'integrating' surveys were very often used by participant. One participant stated: '*We need to create data sets that relate to each other. The more holistic benefits are, the better case for investment it is*' (participant 22, Sustainability). Others mentioned that: '*Aggregation of data is very important to reach conclusions on a global scale'*, and: '*It is about making the data talk to each other and link up'* (participant 18, Transport and Environment; participant 1, Health and Transport). Linking data sets can also reduce survey **costs,** although there were some misconceptions of the advantages of data fusion, for example one participant thought that it would '*increase or boost the sample size'* (participant 8, Health).

As illustrated in Table 4 below, participants described the advantages data fusion could provide, giving practical examples and identifying potential obstacles.

Several suggestions were made regarding which data sets and surveys to combine. According to many, connecting health surveys with some transport questions would be very useful.

Several participants also noticed that the gaps within STATS19 could be solved by cross-referencing STATS19 with HES on the basis of time, data and transport injury (participant 9, Transport and Environment). However, HES does not provide information about people's

Table 4
Data fusion.

Data fusion	
Advantages	• Increase sample size • Reduce costs • Likely to reduce burden on participants
Linking surveys	• Link NTS and HES • Link NTS and Active People Survey • Link Stats 19 and HES • Link Health and Transport Surveys
Practical ideas	• Use the same sample • Add identical questions or aggregate data • Harmonise definitions • Establish common 'identifiers' • Link data from complementary surveys
Obstacles	• Might be burdensome for participants • Might increase costs

general health status, only about the diseases or injuries they have when admitted to hospital. In addition, HES has limitations: official analyses of 'transport injuries' exclude pedestrian falls but HES overestimates travel related cycling injuries (Mindell et al., 2012).

5.2. How can data sets from different surveys be combined?

5.2.1. Participants offered different solutions to enable data fusion, such as using the same sample

Adding *identical questions* to different surveys was offered as a solution to ease 'cross data analysis'. This is easier with harmonised definitions and categories between surveys. The ONS (Office for National Statistics) harmonisation team was created because '*people are becoming more and more aware of data linkage*' (participant 23, Transport, Health and Environment). Participants noted that at the government level, data fusion '*already exists to some extent*' and a '*lot of questions are already harmonised, for example about ethnicity*' (participant 23, Transport, Health and Environment). When not asked of the same participants, this enables information from one source to be linked to different information from a second source for people with identical answers to those identical questions and other characteristics in common.

Aggregating data about the same individuals from different sources was offered as another solution. For example, to link the NTS and HES, you could ask participants in NTS for permission to link their responses to HES. There might be limitations since the *sample will be different* for each survey but '*as long as you are aware of that and as long as you are aware of the limitations the variation in sample sizes can have and you can take that into account and interpret the results*' (participant 8, Health). The participant added: '*We may not be able to make a decision with absolute precision but we are going to be more informed than we would be without it.*' Several participants said that the best solution is to find *common identifiers* or '*intelligent comparators*' (participant 8, Health). Participants felt that surveys should have '*common indices where the datasets can be properly interrelated and agreed upon unified identifiers*' so that surveys can be interrelated through '*certain core attributes*' (e.g. geographic locations, demographics, personal identification) (participant 22, Sustainability). As suggested by one participant, '*the collection of certain indices could be mandatory such as the National Insurance Number or very specific demographics such as date of birth*' (participant 22, Sustainability).

Linking data sets from *complementary sources* was recently used by the NTS team as a way to reduce the survey size, which they had to do for economic reasons. As said by a person in charge of the NTS '*One of the big savings we could make within the questionnaire is we can link data from other data sources so we do not actually have to ask the respondents.*' (participant 23, Transport, Health and Environment). For example, the NTS used to ask participants '*how far it is to their local services';* now they will be using the participant's postcode and link it to Accessibility Indicator data which is done separately (participant 23, Transport, Health and Environment). As a result, '*it reduces the questionnaire length, gives us better quality data and also reduces the burden on the respondents*'. Data fusion might often require getting people's permission to obtain certain data (participant 23, Transport, Health and Environment).

5.3. Platform centralising all surveys: suggestions and barriers

When asked whether establishing an online platform summarising all existing surveys and offering links to surveys would be helpful, many participants agreed that it would, especially in order to gather all the information about surveys which is '*too often spread out*' (participant 14, Transport safety). '*There needs to be a map as to where all these resources are*' added one participant (participant 8, Health). Another person added that there is also a '*Need to invent mechanisms so that people from different fields are aware of the surveys being elaborated*' (participant 22, Sustainability).

Participants suggested that an online platform could '*provide routes to access to surveys*' (participant 2, Transport and Health). In addition to providing a list, this platform could include relevant information or '*insights*' about each survey, '*A paragraph about each survey and an explanation about how each survey relates to each other would be helpful*' (participants 22, Sustainability; participants 1, Health and Transport). The platform could also clearly identify links between surveys and possibilities for data fusion. Examples of existing platforms were given such as: the Public Health Observatories and the National Obesity Observatory (all now part of Public Health

England), UK Data Service, data.gov.uk, ONS, Economics and Social Research Council (ESRC) Question Bank, the Road Safety Observatory, the Road Safety Knowledge Centre, and the European Road Safety Observatory.

6. Discussion

Discussions with stakeholders revealed varying levels of knowledge about the data collected by different surveys and some confusion about operational definitions, such as the recommended physical activity levels according to the Department of Health (see Section 3.1). However, such confusion could be overcome by using a sophisticated online platform as proposed in this paper. This would make surveys more accessible and understandable for users.

Some of the participants' suggestions may be difficult to implement. For example, using the same sample for different surveys would generally be too burdensome for survey participants and would impact on response rates. Some survey questions which were proposed may be too broad – or too specific – to be included in national surveys. Suggested questions such as *"Does cycling improve your well-being",* may need to be more specific to study individuals' perceptions of how cycling affects wellbeing, while separate questions on cycling, wellbeing and confounding factors would be required in a single survey for epidemiological assessment of associations between cycling and wellbeing.

Some participants suggested linking surveys with routine data (e.g. NTS with HES). This would not increase sample size but would increase the range of data available for those survey participants, for far lower costs than additional survey items. There are many examples where data linkage has been achieved and has shed new light on the relationship between transport exposure and health outcomes (e.g. Ward et al. 2002). It also has the advantage of providing subsequent health outcome data, of greater aetiological importance than cross-sectional associations seen in surveys, where reverse causality cannot be excluded (such as greater wellbeing resulting in more cycling, rather than vice versa).

A specific suggestion from several interviewees was for the establishment of a simple, comprehensive online platform describing existing surveys. In theory, this already exists for national surveys in the form of the UK Data Service, which provides comprehensive information about the methods and content of each of these surveys, detailed definitions of the variables, and access to anonymised datasets. Implications of our research include the need for greater publicity about the UK Data Service. However, this does not include the routine, administrative data, so would not provide the desired platform for all relevant datasets.

One of the strengths of this study is the breadth of stakeholders who participated and the many commonalities in points raised by a range of individuals, such as the need for increased data on walking or the need for harmonised definitions across surveys. As a result, the study successfully identified a series of gaps in national data sources which need to be tackled, and several key topics which should be addressed in surveys.

One limitation of the study is that we did not interview sufficient people responsible for organising surveys. Although many participants are very familiar with national surveys and the making of these surveys, they lack the necessary technical understanding to explain the detail of processes such as data fusion. As a result, topics such as the barriers to data fusion could not be discussed at length in the study, yet it is key to improving future surveys (Bleiholder and Naumann, 2009).

Although limited to England or UK data sources, most of the underlying principles will also apply in other countries interested in similar cross-cutting policy issues (such as the implications of encouraging low carbon and active travel modes), particularly those relating to gaps in national surveys (Merom et al., 2010) or methods to address these gaps such as data fusion (Bleiholder and Naumann, 2009; Dong and Naumann, 2009).

7. Conclusion

This study identified a number of gaps in national sources of data (and the clear recognition of the limitations that these gaps caused), the need for new data sets, and proposals for a number of practical solutions. Limitations of current surveys were: insufficient sample size, a lack of harmonisation between surveys, and insufficient links between health, transport, and sustainability. Participants agreed that these deficiencies adversely influence and limit the research, and compromise the ability to evaluate the objectives of national policies. In addition, participants expressed their doubts about the capacity existing surveys have to meet national policy objectives. These limitations could lead to insufficient understanding of how best to meet national policy objectives and the implications of policies under consideration for implementation. This could result, for example, in poor policy decisions with unintended consequences for health and the environment.

The majority of the participants recognised the strong need for joined up research in the fields of transport, health and sustainability. Topics that were highlighted included active travel and its link to health and well-being. More specific themes such as the pleasantness of the environment where people commute were also raised. Many also acknowledged the importance of better representation for certain groups of the population, especially older people and those from deprived areas. When asked for solutions and ideas, participants highlighted the need to harmonise all surveys by establishing common indices across surveys to facilitate data fusion, which was mentioned as being key for the future of national surveys. The ONS data harmonisation programme may go some way addressing this. Other ways forward include increased funding of research aimed at improving data fusion techniques.

Annex 1. Topic Guide

I. Existing data sets

(1) Do you use datasets to explore the relationship between mobility, safety, health, and sustainability?
    a. If yes what data sets do you use?
    b. What are the advantages and disadvantages of current datasets?

    c. How, if at all, are these datasets linked?

(2) To what extent are existing surveys at the National level able to meet the aims & objectives of our National Policies?

(3) To what extent are existing surveys able to analyse the relationship between policies and outcomes?

(4) To what extent do the limitations of existing datasets influence and limit research?

Need for new data sets?

(1) To what extent is there a need for a survey that explores the interrelationship between mobility, health, safety and sustainability?

    a. What gaps in knowledge would such a data set help address?

    b. What questions are you interested in? / What do we need to include in such a data set?

What kind of new data sets are needed?

(1) Possibility to Gather/Combine existing data sources

    a. What should it look like?

    b. What are the possible difficulties/barriers?

    c. How would we overcome them?

    d. What is the best forum to gather all the actors responsible for surveys at the National level

(2) Add new information to existing data sets

    a. Would it be feasible to add data to existing survey?

    b. If yes, how should we do it?

(3) National Survey

    a. What should the ideal survey be like according to you?

    b. What are the possible difficulties/barriers?

    c. How would we overcome them?

    d. To what extent do we need to give priority to obtaining National Data versus Local Data?

References

Bhatia, R., Wier, M., 2011. Safety in Numbers" re-examined: can we make valid or practical inferences from available evidence. Accid. Anal. Prevent. 43, 235–240. http://dx.doi.org/10.1016/j.aap.2010.08.015.

Bleiholder, J., Naumann, F., 2009. Data fusion. ACM Comput. Surv. 41 (1), 41. http://dx.doi.org/10.1145/1456650.1456651 (1–1).

Booth, K.M., Pinkston, M.M., Poston, W.S.C., 2005. Obesity and the built environment. J. Am. Diet. Association 105, 110–117. http://dx.doi.org/10.1016/j.jada.2005.02.045.

Christie, N., Ward, H., Kimberlee, R., Towner, E., Sleney, J., 2007. Understanding high traffic injury risks for children in low socioeconomic areas: a qualitative study of parents' views. Injury Prevent. 13, 394–397. http://dx.doi.org/10.1136/ip.2007.016659.

De Hartog, J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118, 1109–1116. http://dx.doi.org/10.1289/ehp.0901747.

Donaldson, L., 2004. At Least Five a Week: Evidence on the Impact of Physical Activity and its Relationship to Health. Department of Health, London.

Dong, X.L., Naumann, F., 2009. Data fusion: resolving data conflicts for integration. Proc. VLDB Endow 2, 1654–1655.

Dugdill, L., Crone, D., Murphy, R., 2009. Physical Activity and Health Promotion: Evidence-Based Approaches to Practice. John Wiley & Sons, UK.

Edwards, P., 2006. Deaths from injury in children and employment status in family: analysis of trends in class specific death rates. BMJ 333, http://dx.doi.org/10.1136/bmj.38875.757488.4F (119–0).

Graham, D., Glaister, S., Anderson, R., 2005. The effects of area deprivation on the incidence of child and adult pedestrian casualties in England. Accid. Anal. Prevent. 37, 125–135. http://dx.doi.org/10.1016/j.aap.2004.07.002.

Hillman, M., 1993. Cycling and the promotion of health. Policy Stud. Policy Stud. 14, 49–58.

Jarrett, J., Woodcock, J., Griffiths, U.K., Chalabi, Z., Edwards, P., Roberts, I., Haines, A., 2012. Effect of increasing active travel in urban England and Wales on costs to the National Health Service. Lancet 379, 2198–2205. http://dx.doi.org/10.1016/S0140-6736(12)60766-1.

Law, C., Power, C., Graham, H., Merrick, D., 2007. Obesity and health inequalities. Obes. Rev. 8, 19–22. http://dx.doi.org/10.1111/j.1467-789X.2007.00312.x.

Merom, D., van der Ploeg, H.P., Corpuz, G., Bauman, A.E., 2010. Public health perspectives on household travel surveys: active travel between 1997 and 2007. Am. J. Prevent. Med. 39, 113–121. http://dx.doi.org/10.1016/j.amepre.2010.04.007.

Miles, M.B., Huberman, A.M., Saldaña, J., 2014. Qual. Data Anal. Methods Sourcebook.

Mindell, J., Cohen, J.M., Watkins, S., Tyler, N., 2011b. Synergies between low-carbon and healthy transport policies. Proc. Inst. Civil Eng. Transp. 164, 127–139.

Mindell, J., Watkins, S., Cohen, J., 2011a. Health on the move, Policies for Health Promoting Transport. Transport and Health Study Group, Stockport (2).

Mindell, J.S., Leslie, D., Wardlaw, M., 2012. Exposure-based, "like-for-like" assessment of road safety by travel mode Using routine health data. PLoS One 12 (7), e50606. http://dx.doi.org/10.1371/journal.pone.0050606.

National Obesity Observatory, NHS, 2009. Physical activity surveillance in England. What is measured and where are the gaps?.

Srinivasan, S., O'Fallon, L.R., Dearry, A., 2003. Creating healthy communities, healthy homes, healthy people: initiating a research agenda on the built environment and public health. Am. J. Public Health 93, 1446–1450.

Sustrans, 2013. Getting Britain cycling and walking [WWW Document]. URL ⟨http://www.sustrans.org.uk/sites/default/files/images/files/connect2/Getting%20Britain%20Cycling%20and%20Walking%20WEB.pdf⟩.

Transport for London, 2011. Travel in London Report. Transport for London, London (4).

Ward, H., Robertson, S., Lester, T., Pedler, A., 2002. Reporting of road traffic accidents in London: matching police STATS19 with hospital accident and emergency department data. Transport Research Laboratory (London, UK).

Watkins, K., 2012. Safe and Sustainable Roads: An Agenda for Rio+20.

Woodcock, J., Franco, O.H., Orsini, N., Roberts, I., 2011. Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. Int. J. Epidemiol. 40, 121–138. http://dx.doi.org/10.1093/ije/dyq104.

Woodcock, J., Givoni, M., Morgan, A.S., 2013. Health impact modelling of active travel visions for England and Wales using an Integrated Transport and Health Impact Modelling Tool (ITHIM). PLoS One 8, e51462. http://dx.doi.org/10.1371/journal.pone.0051462.