Impacts of study design on sample size, participation bias, and outcome measurement: A case study from bicycling research

Peer-reviewed author version

Branion-Calles, Michael; Winters, Meghan; Nelson, Trisalyn; De Nazelle, Audrey; INT PANIS, Luc; Avila-Palencia, Ione; Anaya-Boig, Esther; Rojas-Rueda, David; DONS, Evi & Götschi, Thomas (2019) Impacts of study design on sample size, participation bias, and outcome measurement: A case study from bicycling research. In: Journal of Transport & Health, 15 (Art N° ARTN 100651).

DOI: 10.1016/j.jth.2019.100651
Handle: http://hdl.handle.net/1942/30354

Impacts of study design on sample size, participation bias, and outcome measurement: a case study from bicycling research

Michael Branion-Calles [a,b], Meghan Winters [a,b], Trisalyn Nelson [c], Audrey de Nazelle [d], Luc Int Panis [e,f], Ione Avila-Palencia [g,h,i,j], Esther Anaya-Boig [d], David Rojas-Rueda [g,k], Evi Dons [e,l], Thomas Götschi [m]

[a] Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

[b] Centre for Hip Health and Mobility, Vancouver, Canada

[c] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, USA

[d] Centre for Environmental Policy, Imperial College London, London, United Kingdom

[e] Flemish Institute for Technological Research (VITO), Mol, Belgium

[f] Transportation Research Institute (IMOB), Hasselt University, Diepenbeek, Belgium

[g] ISGlobal, Barcelona, Spain 4Universitat

[h] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[i] CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

[j] Urban Health Collaborative, Dornsife School of Public Health, Drexel University, Philadelphia, USA

[k] Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, USA

[l] Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium

[m] School of Planning, Public Policy and Management, College of Design, University of Oregon, Eugene, USA

Michael Branion-Calles (Corresponding Author)

Simon Fraser University
Blusson Hall, Room 11300
8888 University Drive
Burnaby, B.C.
V5A 1S6

michael_branion-calles@sfu.ca

Abstract

Introduction: Measuring bicycling behaviour is critical to bicycling research. A common study design question is whether to measure bicycling behaviour once (cross-sectional) or multiple times (longitudinal). The Physical Activity through Sustainable Transport Approaches (PASTA) project is a longitudinal cohort study of over 10,000 participants from seven European cities over two years. We used PASTA data as a case study to investigate how measuring once or multiple times impacted three factors: a) sample size b) participation bias and c) accuracy of bicycling behaviour estimates.

Methods: We compared two scenarios: i) as if only the baseline data were collected (cross-sectional approach) and ii) as if the baseline plus repeat follow-ups were collected (longitudinal approach). We compared each approach in terms of differences in sample size, distribution of sociodemographic characteristics, and bicycling behaviour. In the cross-sectional approach, we measured participants long-term bicycling behaviour by asking for recall of typical weekly habits, while in the longitudinal approach we measured by taking the average of bicycling reported for each 7-day period.

Results: Relative to longitudinal, the cross-sectional approach provided a larger sample size and slightly better representation of certain sociodemographic groups, with worse estimates of long-term bicycling behaviour. The longitudinal approach suffered from participation bias, especially the drop-out of more frequent bicyclists. The cross-sectional approach under-estimated the proportion of the population that bicycled, as it captured ‘typical’ behaviour rather than 7-day recall. The magnitude and directionality of the difference between typical weekly (cross-sectional approach) and the average 7-day recall (longitudinal approach) varied depending on how much bicycling was initially reported.

Conclusions: In our case study we found that measuring bicycling once, resulted in a larger sample with better representation of sociodemographic groups, but different estimates of long-term bicycling behaviour. Passive detection of bicycling through mobile apps could be a solution to the identified issues.

Keywords

Bicycling; Bias; Exposure, Survey participation; Longitudinal; Cross-sectional; Study design

1.0 Introduction

Measuring bicycle behaviour is critical to surveillance of bicycling and its outcomes, including health benefits and crash risks (Götschi et al., 2016). Many population studies rely on indirect measures of bicycle use (e.g., self-reports) as these have low participation burden, are practical to implement, and represent a cost-effective means of collecting a large amount of data (Dishman et al., 2001). As a result, self-report data can facilitate large sample sizes to address myriads of research questions on bicycling behaviour, such as identifying correlates of bicycling or bicycling safety (Kerr et al., 2016; Vanparijs et al., 2015) or quantifying the effect of interventions (Hosford et al., 2018).

Self-reported bicycling can be measured through survey questionnaires or travel diaries (Krizek et al., 2009). These may measure duration or distance of bicycling, or physical activity more broadly (de Geus et al., 2012; Dons et al., 2015; Hosford et al., 2018; Sylvia, 2015). There is no single instrument to measure bicycling behaviour; rather, there are many variations ranging from simple frequency questions to elaborate travel diaries. Instruments may use different units (e.g., time and/or distance) over different time periods (e.g., a day, week or a month)(de Geus et al., 2012; Tin Tin et al., 2013). Furthermore, surveys may be based either on a participants' recall of their bicycling in a specified time period (e.g., in the week prior to the survey) or based on their perception of their average long-term behaviour (e.g., in a "*typical" or "usual"* week). As temporal and seasonal fluctuations are strong for active transportation (Tin Tin et al., 2012; Yang et al., 2011), the timing implied in questions may contribute to variation in bicycling behaviour estimates.

A common study design question in bicycling research and practice is whether to measure participants' bicycling behaviour once (cross-sectional) or multiple times (longitudinal). A cross-sectional approach can be more cost effective with lower burden, enabling wider participation and larger sample sizes. It also does not alter participant's bicycling behaviour. However, given the seasonal variations in bicycling, it may not capture long-term behaviour. Repeated measures, as in a longitudinal study, may provide more accurate measurement of long-term bicycling behaviour as they follow participants through time (including various fluctuations with seasonality, weather, life changes, etc.). This may be especially true for individuals who are sporadic or infrequent bicyclists and may have less accurate recall of their typical behaviours, relative to those that either never bicycle or bicycle routinely (Prince et al., 2008).

1.1 Research Aim

To guide future studies, our aim was to investigate the impacts of study design on the measurement of bicycling behaviour. Specifically, we explored a common question facing both researchers and practitioners: should they collect data once (cross-sectional) or multiple times (longitudinal)?

We capitalized on the Physical Activity through Sustainable Transport Approaches (PASTA) project, a longitudinal cohort study of participants from seven European cities over two years (Dons et al., 2015). We used PASTA data as a case study to investigate how measuring once or multiple times impacted three major factors: a) sample size b) participation bias and c) accuracy of bicycling behaviour estimates. To do so we compared two scenarios: i) as if only the baseline data were collected (the cross-sectional approach) and ii) as if the baseline plus repeat follow-ups were collected (longitudinal approach). The different scenarios, the population samples and analysis approaches for each are outlined in Table 1.

Table 1. Research questions to understand the impacts of study design choices: collecting data once (cross-sectional) or multiple times (longitudinal)

Question	PASTA Subset	Approach
1. Sample size		
1.1 How many participants completed the baseline survey on bicycling compared to subsequent follow-ups?	All PASTA participants that complete the baseline survey (n=7,704).	Total the number of participants that completed baseline self-report and subsequent follow-ups. Calculated the percent change in number of participants (attrition) after each follow-up survey.
2. Participation bias		
2.1. How do geographic, sociodemographic, attitudinal and bicycling behaviour vary between the participants who complete the baseline relative to those that also complete a follow-up?	Participants that complete the baseline survey (n=7,704) versus those that complete at least one follow-up (n=5,806).	Compared geographic, sociodemographic, attitudinal and bicycling behaviour characteristics between each approach using the ratio of relative frequencies. Assessed significant differences through bootstrapped confidence intervals.
2.2. How does the amount of bicycling compare between those who report more follow-ups relative to those that complete less?	Participants that complete at least one follow-up and report some bicycling (n =3,511).	Calculated each participant's average 7-day bicycling behaviour in minutes over all follow-ups completed. Modeled participants' average 7-day minutes of bicycling as a function of the number of follow-ups they completed using a GAM.
3. Accuracy of bicycling behaviour estimates		
3.1 Is binary bicycling behaviour (yes or no) consistent from baseline to follow-ups?	Participants in the longitudinal study (n=5,806).	Categorized participants' bicycling status (yes/no) at baseline, and over each follow-up. Generated a confusion matrix for bicycling status.

3.2. Are bicycling behaviour estimates similar when calculated from i) typical weekly bicycling at baseline and ii) repeated measures of bicycling in last 7 days?	Participants who provided non-zero estimates of bicycling duration at baseline and who completed at least 1 follow-up (n = 2,635)	Modeled the absolute difference between average follow-up bicycling behaviour and baseline typical bicycling behaviour using a GAM to understand magnitude and directionality of errors.

GAM = Generalized Additive Model.

2.0 Materials and Methods

2.1 Study Design

Data were collected as part of the PASTA project, funded by the EC under FP7-HEALTH-2013-INNOVATION-1. Data from a longitudinal web-based survey of residents of Antwerp, Barcelona, London, Örebro, Rome, Vienna and Zürich (Dons et al., 2015) were collected between November 2014 (April 2015 in Örebro) and December 2016. Participants could enter the study at any time. The study employed an opportunistic sampling approach, although a portion of participants in Örebro were recruited through random sampling. The same standards for recruitment were used in all cities, including press releases and editorials, integrated promotional materials, collaboration with local stakeholders networks to distribute information, promotion of the study through social media and participation incentivization though a prize lottery (except for Örebro where lotteries were not legally permitted) (Gaupp-Berghausen et al., 2019). All promotional materials and automated questionnaires were translated into local languages by native speakers. A custom survey platform sent up to three automatic reminder emails to complete questionnaires. Participants were 18 years or older, except for in Zürich, where the minimum age was 16 years. Bicyclists were oversampled in order to have sufficient samples in cities with a low bicycling mode share (Raser et al., 2018).

The surveys consisted of a comprehensive baseline questionnaire followed by repeated frequent short and long follow-up surveys (Figure 1). The baseline questionnaire collected data on sociodemographic characteristics, travel behaviour, physical activity, locational information (home, work and school), as well as attitudes toward transportation. Physical activity questions included a modified version of the Global Physical Activity Questionnaire (GPAQ) aimed at estimating the duration and frequency of bicycling (Gerike et al., 2016). The entire baseline survey was designed to take 30 minutes to complete (Dons et al., 2015). Following the baseline survey, a short follow-up survey was sent out every 13 days to collect measurements of physical activity and travel behaviour in the previous 7 days. This was designed to take 5 minutes to complete (Dons et al., 2015). A long follow-up survey was sent out every third follow-up; which was identical to the short follow-up but with the addition of a 1-day travel diary. The long follow-up was designed to take 10 minutes to complete (Dons et al., 2015). At each follow-up, participants were also given the opportunity to report any safety incidents (e.g., crashes) they experienced since their last follow-up.

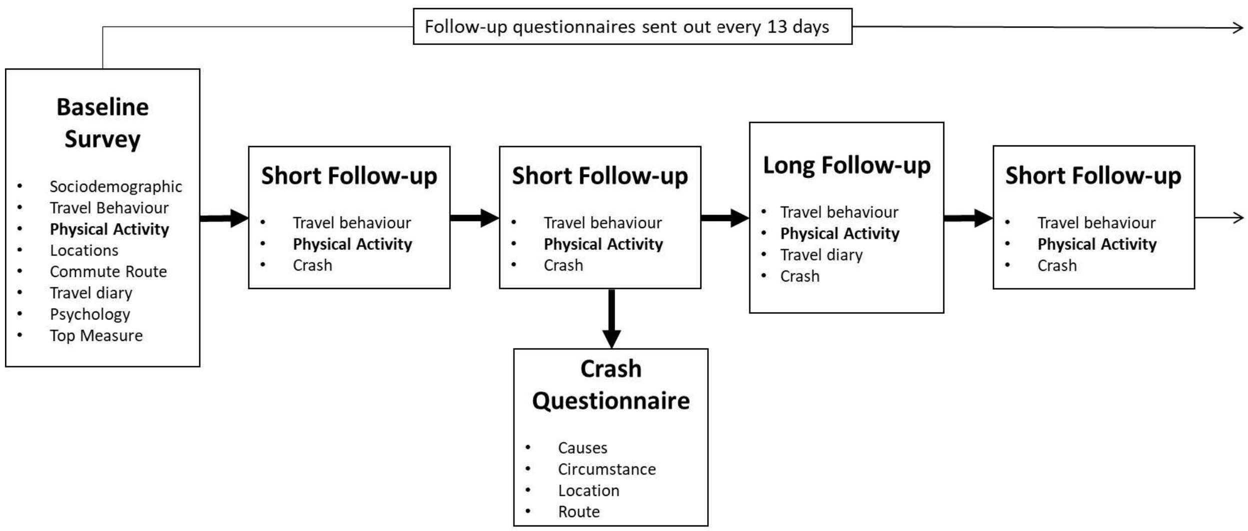

Figure 1. PASTA study design.

In the baseline questionnaire the modified version of the GPAQ asked two questions to estimate long-term bicycling behaviour: 1) "In a *typical* week, on how many days do you cycle for at least 10 minutes continuously to get to and from places? and 2) "Typically, how much time do you spend cycling on such a day?" The same questions were asked for each follow-up survey, but the time period was framed as the prior seven days, rather than for a typical week.

2.2 Data processing and cleaning

Typical and average 7-day recall were calculated for each participant. Typical weekly bicycling was calculated by multiplying the number of days they typically bicycle by the time spent bicycling on those days. Average 7-day recall was estimated by first calculating the time spent bicycling in previous 7-days for each follow-up, and then taking the average over follow-ups.

We removed all participants affected by a proposed intervention ("top measures") within the broader PASTA project, as survey administration differed for this group. These participants were identified a priori as "exposed" to an urban form change or participation in a program within the study period, and were placed into a "hibernation period" before the planned intervention, in which they were not sent new questionnaires (Dons et al., 2015).

We then defined the two study design approaches using the PASTA study: cross-sectional and longitudinal. In the cross-sectional approach, we only considered a participant's baseline-questionnaire, while in the longitudinal approach we considered their follow-ups. The participants within the cross-sectional approach consisted of those that completed the GPAQ component of the baseline questionnaire and did not provide outlier values. Outlier values were defined as bicycling >8 hours on a given day in a typical week. The participants within the longitudinal approach consisted of the subset from the cross-sectional approach which completed the GPAQ component of at least 1 follow-up survey and did not provide outlier values in any of their follow-ups. Outlier values in the longitudinal approach were defined as reporting bicycling an average of >8 hours on a given day in the past week. A flowchart of the process is presented in Figure 2.

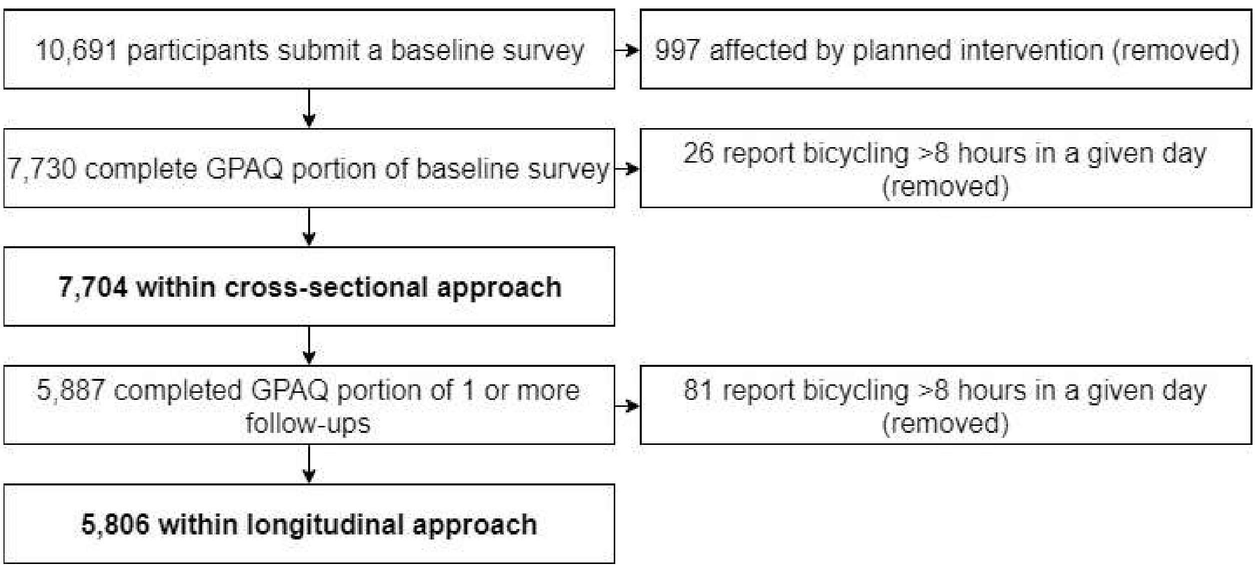

Figure 2. Data cleaning flow-chart to define two study design approaches: the cross-sectional and longitudinal approach.

2.3 Analysis

2.3.1 Sample Size

To understand the impact of measuring once versus multiple times on sample size we compared the number of participants who completed baseline self-report to the number who completed subsequent follow-ups (Table 1, Question 1.1). We also calculated the percent change in number of participants after each follow-up survey to understand patterns of attrition. The number of participants who completed the baseline survey represents the sample size for the cross-sectional approach, while the number of participants who completed at least the first follow-up represents the sample size for the longitudinal approach.

2.3.2 Participation Bias

We compared the relative frequencies of sociodemographic, attitudinal and bicycling characteristics at baseline between the cross-sectional and longitudinal approaches. Sociodemographic characteristics we included were age, gender, body mass index, education, income, employment, drivers licensing and having young children. Attitudinal characteristics included the participants level of comfort, and perceived safety of bicycling for transport, as well as how well regarded and common they felt bicycling was in their neighbourhood. Bicycling characteristics included the frequency of bicycling at baseline and whether they typically bicycled in a given a week. We compared the ratio of relative frequencies (RRF) between each level of a given variable of interest (longitudinal approach / cross sectional approach) (Table 1, Question 2.1) (Tin Tin et al., 2014). An RRF of 1 corresponds to no change in representation of given characteristic from a cross-sectional to longitudinal approach, while > 1 corresponds to over-representation and <1 under-representation. We constructed a 95% confidence interval around each RRF through bootstrapping with 10,000 replications to assess statistical significance (Tin Tin et al., 2014).

Participants within the longitudinal approach completed varying numbers of follow-ups, so we sought to understand if there was an association between the number of follow-ups completed and the average 7-day recall over those follow-ups (Table1, Question 2.2). To do so, we modeled participants' average 7-day recall (average over all follow-ups) as a function of the number of follow-ups they completed. We restricted this analysis to the subset of participants within the longitudinal approach (i.e., the participants with repeat measurements) who reported some bicycling and considered up to the first 28 follow-up surveys completed (~ 1 year of follow-ups if completed every 13 days). We used a generalised additive model (GAM) with thin-plate spines to estimate the shape of the relationship between participants' overall average 7-day recall and their number of completed follow-ups.

2.3.3 Accuracy of Bicycling Behaviour Estimates

To assess whether accuracy of bicycling status was consistent from baseline to follow-ups, we compared bicycling status derived from typical weekly bicycling to bicycling status from average 7-day recall. We only considered the first 28 follow-ups in calculating average 7-day recall (~ 1 year of follow-ups if completed every 13 days) (Table 1, Question 3.1). Participants were coded as "typical bicyclists" if they provided non-zero values for bicycling duration in a typical week at baseline. They were coded as "follow-up bicyclists" if they had non-zero values for bicycling in the previous 7 days in any follow-up. We assess consistency of bicycling status between baseline and follow-up by framing this as 'false negative' and 'false positive rates'. In this instance, the false negative rate refers to the proportion of participants who bicycle in follow-ups that were not identified as bicyclists at baseline, while the false positive rate refers to the proportion of participants who were identified as bicyclists at baseline but reported no bicycling in follow-ups.

One-time surveys often ask participants to recall their typical bicycling habits over a period of time to estimate long-term average behaviour. In contrast, when there are repeated measurements researchers may use the averaged value to estimate long-term behaviour. Thus, we sought to understand if estimates of typical weekly bicycling at baseline were similar to average 7-day recall reported over follow-up surveys, quantifying the absolute and relative differences between them (Table 1, Question 3.2). We only considered up to 28 follow-up surveys (~ 1 year of follow-ups if completed every 13 days). For each participant we calculated absolute error by subtracting their typical weekly bicycling at baseline from their average 7-day recall over follow-ups. The shape of the relationship between typical weekly bicycling at baseline and the absolute error was estimated using a generalised additive model with thin-plate splines (Zuur et al., 2009).

We visualised the differences between typical weekly bicycling at baseline and average 7-day recall over follow-ups by developing a correction factor (typical weekly bicycling/ the predicted average 7-day recall) for the range of typical weekly bicycling values. Values above 1 indicate the need to correct for under-predictions at baseline and below 1, overpredictions. Since the number of follow-ups may affect the accuracy, we also examined the relationship between number of completed follow-ups and the absolute error.

3.0 Results

3.1 Sample Size

There were 10,691 participants who submitted a baseline survey but only 7,704 of these completed the GPAQ component. These participants made up the participants within the cross-sectional approach (Figure 3a). Of the participants in the cross-sectional approach 5,806 participants completed the GPAQ component of at least the first follow-up survey and comprise the participants within the longitudinal approach. This represents an attrition of 24.6% from baseline to the first follow-up (Figure 3b). The attrition rate was highest in the initial follow-ups (10.4% - 16.1% attrition over follow-ups 2-4) and lessened later on (4.3% - 10.8% attrition from follow-up 5-35). Only a small proportion of participants completed 36 or more follow-ups (4.7%), meaning there were larger relative incremental percentage change in sample size in the later follow-ups. Because of rolling recruitment, participants would have needed to have been in the study for over a year to complete more than 30 surveys.

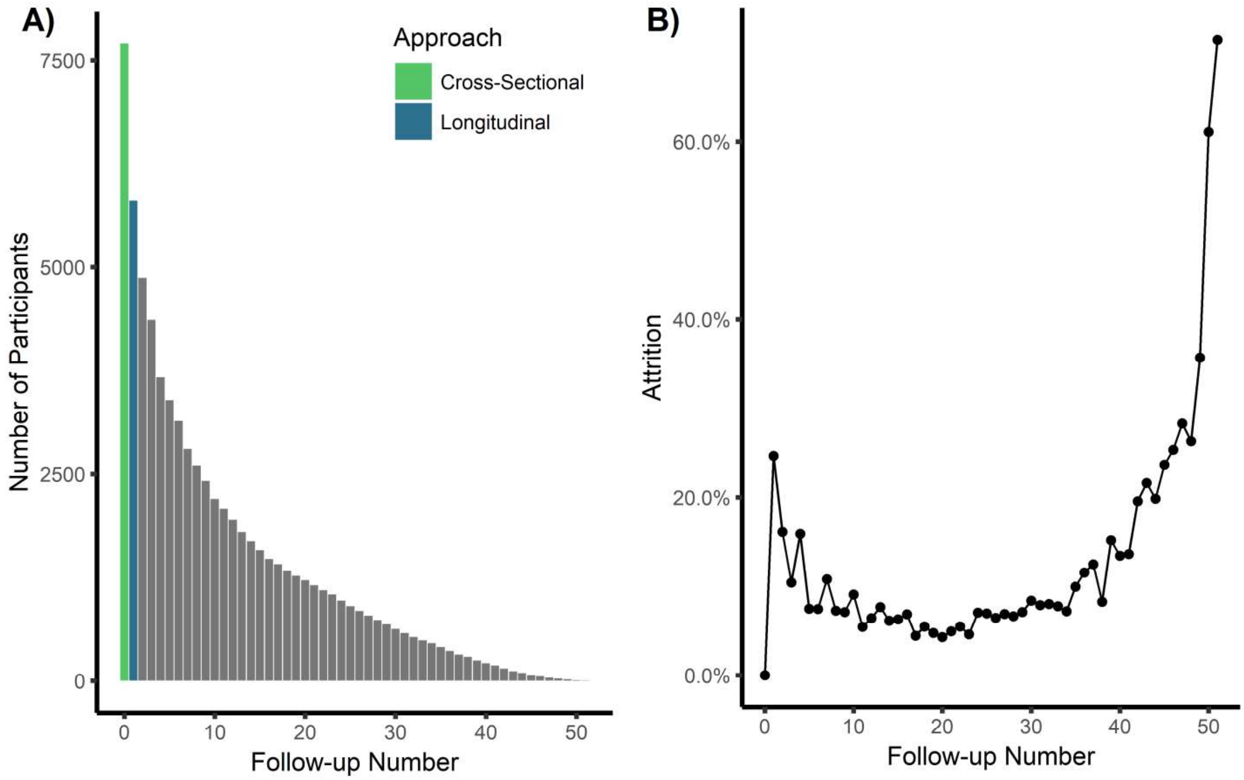

Figure 3. A) The cumulative number of participants completing GPAQ follow-up surveys. The green column represents participants who, at minimum, complete the GPAQ component of the baseline and comprise the "baseline approach"; the blue those who, at minimum, completed the first follow-up survey and comprise the "longitudinal approach". B) The attrition in total number of participants at each follow-up survey. For example, 24.6% of participants did not complete the first follow-up after the baseline, while 16.1% do not complete the second follow-up after the first.

3.2 Participation Bias

3.2.1 How do geographic, sociodemographic, attitudinal and bicycling behaviour characteristics vary between the participants who complete the baseline relative to those that also complete a follow-up?

There were differences in the distribution of geographic and sociodemographic characteristics of participants within the longitudinal approach relative to the cross-sectional. Residents of Zürich were over-represented, while residents of London and Örebro were under-represented (Table 3). Sociodemographic groups that were slightly over-represented in the longitudinal approachincluded those with a normal BMI, the highly educated, middle-income, and those without children 18 years or under. Slightly under-represented groups included students. Participants aged 16-25 years or over 65+ years were also less likely to be within the longitudinal approach. The longitudinal approach had much lower rates of missing data for some sociodemographic characteristics including BMI, education, income, having young children, and perceptions of bicycling in their neighbourhood.

Table 2. Sociodemographic, attitudinal and bicycling characteristics of participants by participation

		Cross-Sectional		Longitudinal		
Variable	Level	Frequency	Relative Frequency	Frequency	Relative Frequency	RRF (95 % CI)
n		7704		5806		
City	Antwerp	884	11.5	705	12.1	1.06 (0.96, 1.16)
	Barcelona	1400	18.2	1073	18.5	1.02 (0.95, 1.09)
	London	1074	13.9	715	12.3	**0.88 (0.81, 0.96)**
	Örebro	560	7.3	355	6.1	**0.84 (0.74, 0.96)**
	Rome	1512	19.6	1087	18.7	0.95 (0.89, 1.02)
	Vienna	1132	14.7	896	15.4	1.05 (0.97, 1.14)
	Zürich	1142	14.8	975	16.8	**1.13 (1.05, 1.22)**
Age (years)	16-25	1186	15.4	826	14.2	0.92 (0.85, 1.00)
	26-35	2301	29.9	1731	29.8	1.00 (0.95, 1.05)
	36-45	1816	23.6	1401	24.1	1.02 (0.96, 1.09)
	46-55	1485	19.3	1153	19.9	1.03 (0.96, 1.10)
	56-65	666	8.6	528	9.1	1.05 (0.94, 1.17)
	65+	248	3.2	165	2.8	0.88 (0.72, 1.07)
	Missing	2	0.0	2	0	1.33 (0.00, 5.31)
Gender	Female	4061	52.7	3073	52.9	1.00 (0.97, 1.04)
	Male	3643	47.3	2733	47.1	1.00 (0.96, 1.03)
BMI	<25	5197	67.5	4044	69.7	**1.03 (1.01, 1.06)**
	25-30	1741	22.6	1315	22.6	1.00 (0.94, 1.07)
	30+	547	7.1	395	6.8	0.96 (0.84, 1.09)
	Missing	219	2.8	52	0.9	**0.32 (0.23, 0.42)**
Education	No degree	24	0.3	11	0.2	0.61 (0.27, 1.22)
	Primary education	93	1.2	67	1.2	0.96 (0.69, 1.30)
	Secondary/further education	2006	26.0	1498	25.8	0.99 (0.94, 1.05)
	Higher/university education	5320	69.1	4200	72.3	**1.05 (1.02, 1.07)**
	Missing	261	3.4	30	0.5	**0.15 (0.10, 0.21)**
Income (€)	<10,000	711	9.2	492	8.5	0.92 (0.82, 1.02)
	10,000 - 24,999	1222	15.9	937	16.1	1.02 (0.94, 1.10)
	25,000 - 49,999	1837	23.8	1473	25.4	**1.06 (1.00, 1.13)**
	50,000 - 74,999	1150	14.9	950	16.4	**1.10 (1.01, 1.19)**
	75,000 - 99,999	527	6.8	413	7.1	1.04 (0.92, 1.18)
	100,000 - 150,000	291	3.8	251	4.3	1.14 (0.97, 1.35)
	>150,000	113	1.5	90	1.60	1.06 (0.80, 1.39)
	Missing	1853	24.1	1200	20.7	**0.86 (0.81, 0.92)**

Employment	Full-time employed	4437	57.6	3410	58.7	1.02 (0.99, 1.05)
	Part-time employed, or casual work	1280	16.6	1021	17.6	1.06 (0.98, 1.14)
	Student / In training	1142	14.8	790	13.6	**0.92 (0.84, 1.00)**
	Home duties / Unemployed / Retired / Sickness leave / Parental leave	661	8.6	462	8	0.93 (0.83, 1.04)
	Missing	184	2.4	123	2.1	0.89 (0.70, 1.11)
Has Driver's License	Yes	6737	87.4	5128	88.3	1.01 (1.00, 1.02)
	No	967	12.6	678	11.7	0.93 (0.85, 1.02)
Has Children Under 18 years	Yes	2452	31.8	1884	32.4	1.02 (0.97, 1.07)
	No	4715	61.2	3684	63.5	**1.04 (1.01, 1.06)**
	Missing	537	7.0	238	4.1	**0.59 (0.50, 0.68)**
Bicycling for transport is comfortable	Agree	4398	57.1	3369	58	1.02 (0.99, 1.05)
	Neutral	1715	22.3	1262	21.7	0.98 (0.92, 1.04)
	Disagree	1591	20.7	1175	20.2	0.98 (0.92, 1.05)
Bicycling for transport is safe with regards to traffic	Agree	1586	20.6	1165	20.1	0.97 (0.91, 1.04)
	Neutral	1779	23.1	1343	23.1	1.00 (0.94, 1.06)
	Disagree	4339	56.3	3298	56.8	1.01 (0.98, 1.04)
In my neighbourhood bicycling is well regarded	Agree	3327	43.2	2564	44.2	1.02 (0.98, 1.06)
	Neutral	2605	33.8	2010	34.6	1.02 (0.98, 1.07)
	Disagree	1606	20.8	1232	21.2	1.02 (0.95, 1.09)
	Missing	166	2.2	0	0	
In my neighbourhood bicycling is common	Agree	2646	34.3	2040	35.1	1.02 (0.98, 1.07)
	Neutral	2340	30.4	1801	31	1.02 (0.97, 1.07)
	Disagree	2517	32.7	1965	33.8	1.04 (0.99, 1.09)
	Missing	201	2.6	0	0	
Typical Bicycling	Never	1903	24.7	1365	23.5	0.95 (0.89, 1.01)
	< once per month	1044	13.6	782	13.5	0.99 (0.91, 1.08)
	1-3 days per month	760	9.9	571	9.8	1.00 (0.90, 1.11)
	1-3 days per week	1233	16.0	935	16.1	1.01 (0.93, 1.09)
	Daily or almost daily	2711	35.2	2122	36.5	1.04 (0.99, 1.09)
	Missing	53	0.7	31	0.5	0.78 (0.48, 1.19)
Baseline weekly bicyclist	Yes	3461	44.9	2692	46.4	1.03 (0.99, 1.07)
	No	4243	55.1	3114	53.6	0.97 (0.94, 1.00)

[a] 95% bootstrapped confidence intervals with 10,000 replications.
RRF = Ratio of Relative Frequencies
Bold = statistical significance at 95% confidence.

3.2.2 How does the amount of bicycling compare amongst those who report more follow-ups relative to those that complete less?

Participants with the fewest follow-ups tended to report more minutes of bicycling in their 7-day recall (Figure 4). The predicted average 7-day recall was just over 210 minutes for bicyclists who completed one follow-up, compared to 135 minutes/week for those who completed 15 follow-ups: a 75-minute difference.

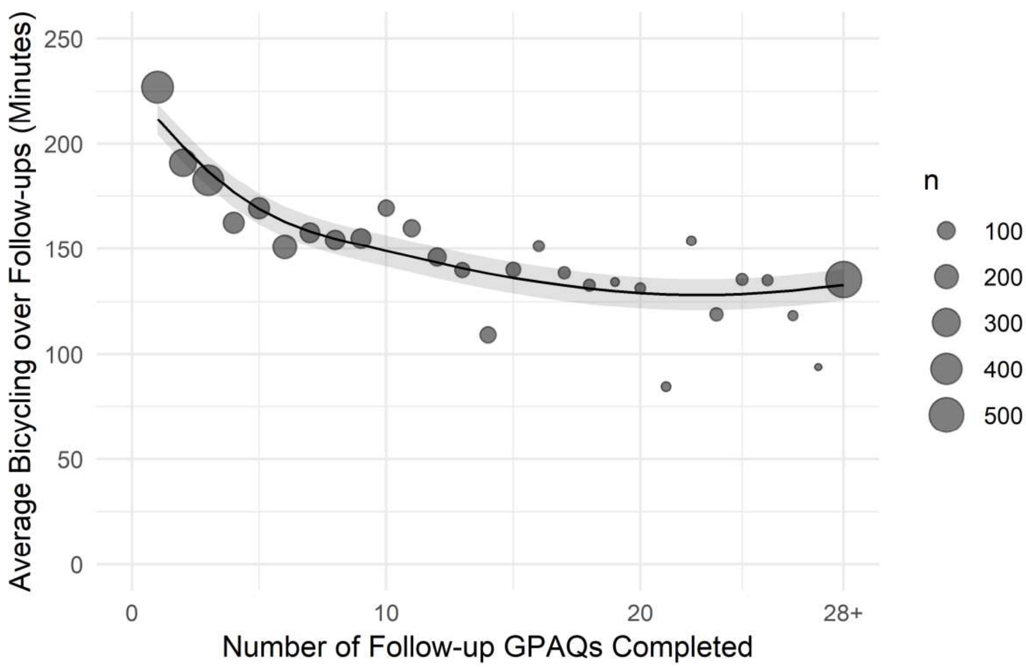

Figure 4. The relationship between the average 7-day recall over follow-ups amongst bicyclists and the number of follow-ups completed. Fitted trend line on the raw data (not plotted) using a simple generalized additive model.

3.3 Accuracy of Bicycling Behaviour Estimates

3.3.1 Are behaviour estimates consistent from baseline to follow-ups?

At baseline 46.4% (2,692 / 5,806) of participants were classified as typical bicyclists, while over follow-ups 60.5% (3,511/5,806) were classified as follow-up bicyclists (Table 4). Typical bicycling status at baseline was consistent with follow-up bicycling status for just over 4 in 5 participants (4,705/5,806). There was a small chance that if participant was coded as a follow-up non-bicyclist, that they previously reported being a typical bicyclist (6.1% false positive rate). There was a comparatively higher chance that if a participant reported being a follow-up bicyclist, that they previously reported being a typical non-bicyclist (27.3% false negative rate).

Table 3. Confusion matrix for bicycling status at baseline (cross-sectional approach) or over follow-ups (longitudinal approach).

		7-Day Recall Over Follow-ups (Up to 28)	

		Follow-up Bicyclist	Follow-up Non-Bicyclist	Total
Baseline Typical Weekly Bicycling (cross-sectional)	Typical Bicyclist	2551 (72.7%)	141 (6.1%)	2692
	Typical Non-Bicyclist	960 (27.3%)	2154 (93.9%)	3114
	Total	3511(100%)	2295 (100%)	5806

3.3.2 Are bicycling behaviour estimates similar when calculated from i) typical weekly bicycling at baseline and ii) repeated measures of bicycling in last 7 days?

There were 5,806 participants who provided duration data on bicycling behaviour in both baseline and follow-ups. For this analysis we considered only the 2,692 participants who were coded as a typical bicyclist at baseline and removed 57 participants that reported typically bicycling more than 2 hours daily.

We found that the accuracy of the typical weekly bicycling estimate at baseline varied by how much bicycling was initially reported, as well as based on the number of follow-up surveys a participant completed. Participants who reported bicycling less than 1.5 hours in a typical week at baseline (~13 minutes a day) tended to report higher levels of bicycling in follow-ups (Figure 5a). There was non-linearity in the relationship between typical bicycling at baseline and the average 7-day recall, with greater over-estimation for participants with higher reported typical weekly bicycling at baseline (Figure 5a). We also found that the number of follow-up surveys completed had a small but significant association with the accuracy of the typical bicycling estimate (Figure 5b). The relationship was linear, where the over-estimation at baseline was increased by just under a minute for every follow-up completed, from a 49-minute weekly overestimation for participants who completed 1 follow-up, increasing to 71-minutes for participations who completed 28 follow-ups.

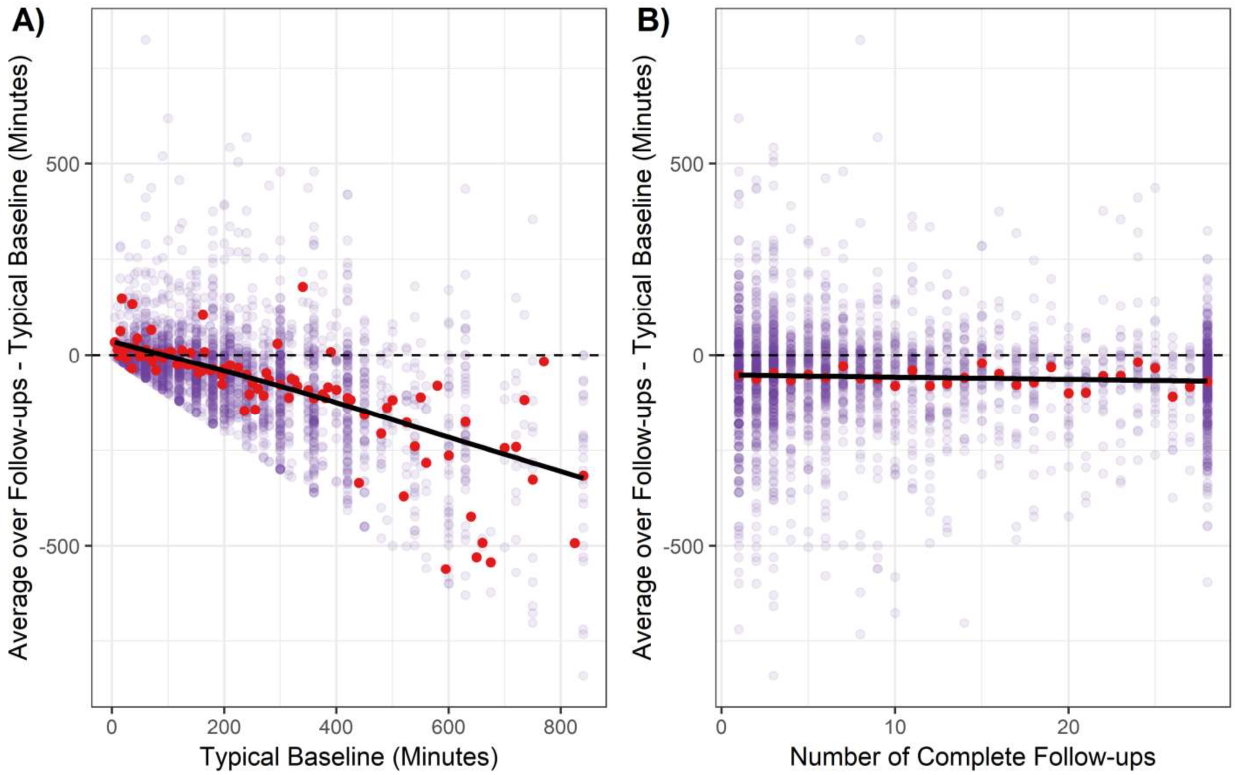

Figure 5. A) The relationship between typical 7-day bicycling measured at baseline and the difference between average 7-day recall over follow-up surveys (1 or more) and typical weekly of bicycling at baseline. B) The relationship between the number of follow-ups and the difference between the average 7-day recall and typical weekly bicycling at baseline. Points above the dotted-black line indicate an under-estimation of minutes of bicycling at baseline, while points below indicate an over-estimation. Red points indicate the mean difference for a given baseline value or number of follow-ups completed. A generalized additive model was used to visualise the trend in the data.

The relative difference between the typical weekly bicycling and average 7-day recall indicate that correction factors decrease non-linearly from 4.2 to 0.6 for typical weekly baseline bicycling values between 10 and 840 minutes (Figure 6). The non-linear decrease can be illustrated through the following hypothetical example: if 6 participants report that they bicycle 10, 30, 60, 240 and 600 minutes in a typical week respectively, the model suggests that the first 3 participants under-predict their average 7-day recall by factors of 4.2, 1.8, and 1.2, while the last 3 participants would over-predict their average 7-day recall by factors of 0.8, 0.7 and 0.6.

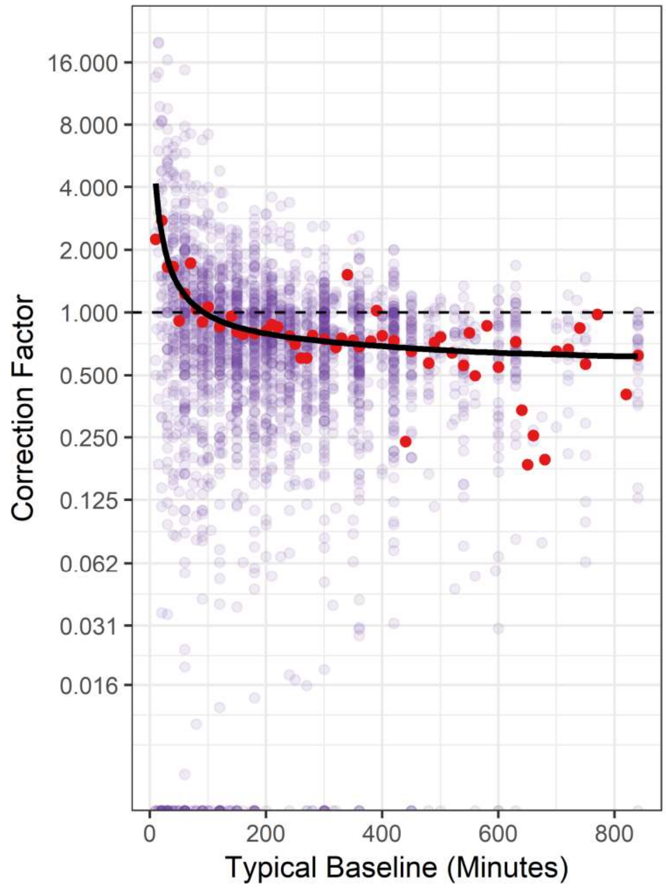

Figure 6. The predicted factor for converting baseline typical bicycling values to the average 7-day recall over follow-ups (black line). A corrective factor above 1 indicates a baseline under-estimation, below 1 an over-estimation. Purple points represent the data, red points the average for a given baseline typical bicycling value.

4.0 Discussion

In this study we used a large longitudinal study with over 10,000 participants in seven European cities to understand the impacts of two study designs (cross-sectional vs longitudinal approaches) on sample size, participation bias and accuracy of bicycling behavior estimates. We found that a cross-sectional approach resulted in a larger overall sample size, and slightly better representation of sociodemographic groups, but inconsistent estimates of long-term bicycling behaviour. In contrast, the longitudinal approach may provide more accurate bicycling behaviour estimates, but suffers from some participation bias, especially the selective drop-out of more frequent bicyclists with greater numbers of follow-up surveys.

Measuring bicycling behaviour accurately is essential for both research and practice. Many studies differentiate between bicyclists and non-bicyclists through self report (Krizek et al., 2009). In a cross-sectional study, differentiating between bicyclists and non-bicyclists can involve dichotomizing participants based on a question that asks for typical or usual bicycling habits within a given time frame (e.g. a week) (Moudon et al., 2005; Winters et al., 2007). To separate participants into bicyclists versus non-bicyclists, our analysis suggests that asking for typical weekly bicycling habits will result in the misclassification of ~1 in 20 bicyclists and ~1 in 4 non-bicyclists. The inconsistency we found could be due to participants having genuinely changed their bicycling behaviour; however, this is unlikely given the short duration of study participation (median time between baseline and follow-up < 5 months for this subset). We suggest it was more likely that the wording of the question itself resulted in the classification issue: participants who may not bicycle in a "typical week" may bicycle in the 7-day recall periods in follow-ups. Questions that ask for direct recall of bicycling for a longer period of time (e.g., in the past 12 months) or use categories (e.g., never, daily, 1-3 days per week, 1-3 days per month, <once per month, etc.) may have better consistency.

We also found that the duration of bicycling derived from self-reported typical weekly bicycling habits was inconsistent with that derived from recall of the past 7-days. When we compared the typical weekly bicycling at baseline to average 7-day recall over follow-ups, we found that bicyclists who reported they typically bicycle frequently (> 90 minutes a week), over-estimated their habits, and those who reported typically bicycling more infrequently (< 90 minutes a week) under-estimated bicycling. Over-estimation is common in self report physical activity as a result of social desirability bias, or recall bias (Brenner and DeLamater, 2014; Dishman et al., 2001; Panter et al., 2014; Sallis and Saelens, 2000). Few studies have assessed measurement validity for bicycling. One study of 11 bicyclists in the UK compared average trip durations derived from GPS data to a questionnaire asking for the "usual" time spent on a bicycling trip and found a mean difference of ~1-minute, and generally good agreement between the methods (Panter et al., 2014). Small errors in durations derived from recall of usual habits at the trip level, however, may compound given aggregation to a weekly time period (Panter et al., 2014).

The implications of the use of typical weekly bicycling to estimate the amount of bicycling in a cross-sectional approach would depend on the population being sampled. For example, consider a cross-sectional study that sought to quantify population crash rates by asking participants to recall prior crashes (numerator) and assessed bicycling through a question regarding their typical bicycling habits (denominator). If crashes were distributed equally, and the sample consisted of a larger proportion of infrequent bicyclists relative to frequent bicyclists, we would over-estimate overall crash risk due to the under-estimation of bicycling for infrequent bicyclists. Conversely, a sample with a greater proportion of more frequent bicyclists relative to infrequent bicyclists would result in an under-estimation of crash risks, with an over-estimation of bicycling amongst frequent bicyclists.

Loss to follow-up is a concern for cohort studies, given the potential impacts for biased associations (Greenland, 1977; Kristman et al., 2004; Tin Tin et al., 2014) if both exposure and outcome are related to study participation (Lash et al., 2009). Our results suggest that there are only slight differences between a select few sociodemographic variables from baseline to the first follow-up, such as people with higher educations, students, middle income earners and people with young children. However, the loss to follow-up did impact bicycling behaviours: we saw a ~75-minute drop in bicycling reported in average 7-day recall for those with 1 follow-up survey, relative to those who completed 15, suggesting a participation bias effect. An alternative explanation for the decrease in bicycling was that it was a short-term effect caused by participation in the study itself (Dishman et al., 2001). We explored this possibility in a separate analyses by plotting the average 7-day recall after each follow-up, for a subset of participants who completed at least 15 follow-ups. The average bicycling within the first follow-up was 149 minutes, while the average bicycling within the fifteenth follow-up was 130 minutes, suggesting a short-term study effect was not substantial. In the PASTA study, participants were also asked to complete a detailed 1-day travel diary at every third follow-up (Gerike et al., 2016). As such there was differential burden for participants who took more trips. The detailed 1-day travel diary would incur a higher burden on participants with many trips (bicycling and other modes) and potentially lead to increased drop out amongst these participants. We expect that in a similar study which does not include a trip diary, the bias may not be as strong.

The PASTA study is one of the largest mobility studies of its kind, and provided a large sample of longitudinal survey data across seven geographically diverse cities in Europe. While we frame the baseline survey as a cross-sectional sample, PASTA respondents were aware they were signing up for a longitudinal survey and may not be completely representative of an independent cross-sectional sample. The survey structure may have impacted answer quality and quantity, as the PASTA baseline survey was long, and follow-ups were frequent (every 13 days). We used the average 7-day recall to assess the accuracy of typical weekly bicycling, but we did not assess the accuracy of the average 7-day recall itself. The GPAQ has been validated against direct measures of physical activity (Bull et al., 2009; Laeremans et al., 2017) but the bicycling-specific questions have not been validated. In estimating participation bias, we only compared changes after the first follow-up and a higher threshold may result in different patterns.

5.0 Conclusions

Future studies aiming to derive measures of bicycling behaviour based on repeated measurements must consider the trade-offs between estimating individual bicycling behaviour more accurately, with bias and power. In our case study we found that measuring bicycling once, compared to multiple times, resulted in a larger sample with better representation of sociodemographic groups and bicyclists, but substantially different estimates of long-term bicycling behaviour. We suggest that measuring typical weekly habits at one point in time is not an accurate proxy for measuring bicycling in the past 7-days multiple times. Problems with participation bias and sample size could be resolved in future studies through the use of app-based studies to capture bicycling behaviour (Geurs et al., 2015), which, if automated and passively collected over time, may one day enable rich travel data at a lower burden to participants than traditional methods (Prelipcean et al., 2017). Further developments are needed for accurate mode detection and privacy considerations (Geurs et al., 2015).

6.0 References

Brenner, P.S., DeLamater, J.D., 2014. Social Desirability Bias in Self-reports of Physical Activity: Is an Exercise Identity the Culprit? Soc. Indic. Res. 117, 489–504. doi:10.1007/s11205-013-0359-y

Bull, F.C., Maslin, T.S., Armstrong, T., 2009. Global Physical Activity Questionnaire (GPAQ): Nine Country Reliability and Validity Study. J. Phys. Act. Heal. 6, 790–804. doi:10.1007/978-3-319-28114-8_7

de Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2012. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accid. Anal. Prev. 45, 683–693. doi:10.1016/j.aap.2011.09.045

Dishman, R.K., Washburn, R.A., Schoeller, D.A., 2001. Measurement of physical activity. Quest 53, 295–309. doi:10.1080/00336297.2001.10491746

Dons, E., Götschi, T., Nieuwenhuijsen, M., de Nazelle, A., Anaya, E., Avila-Palencia, I., Brand, C., Cole-Hunter, T., Gaupp-Berghausen, M., Kahlmeier, S., Laeremans, M., Mueller, N., Orjuela, J.P., Raser, E., Rojas-Rueda, D., Standaert, A., Stigell, E., Uhlmann, T., Gerike, R., Int Panis, L., 2015. Physical Activity through Sustainable Transport Approaches (PASTA): protocol for a multi-centre, longitudinal study. BMC Public Health 15, 1126. doi:10.1186/s12889-015-2453-3

Gaupp-Berghausen, M., Raser, E., Anaya-Boig, E., Avila-Palencia, I., de Nazelle, A., Dons, E., Franzen, H., Gerike, R., Gotschi, T., Iacorossi, F., Hossinger, R., Nieuwenhuijsen, M., Rojas-Rueda, D., Sanchez, J., Smeds, E., Deforth, M., Standaert, A., Stigell, E., Cole-Hunter, T., Int Panis, L., 2019. Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project. J. Med. Internet Res. 21, e11492. doi:10.2196/11492

Gerike, R., de Nazelle, A., Nieuwenhuijsen, M., Panis, L.I., Anaya, E., Avila-Palencia, I., Boschetti, F., Brand, C., Cole-Hunter, T., Dons, E., Eriksson, U., Gaupp-Berghausen, M., Kahlmeier, S., Laeremans, M., Mueller, N., Orjuela, J.P., Racioppi, F., Raser, E., Rojas-Rueda, D., Schweizer, C., Standaert, A., Uhlmann, T., Wegener, S., Götschi, T., 2016. Physical Activity through Sustainable Transport Approaches (PASTA): a study protocol for a multicentre project. BMJ Open 6, e009924. doi:10.1136/bmjopen-2015-009924

Geurs, K.T., Thomas, T., Bijlsma, M., Douhou, S., 2015. Automatic trip and mode detection with move smarter: First results from the Dutch Mobile Mobility Panel. Transp. Res. Procedia 11, 247–262. doi:10.1016/j.trpro.2015.12.022

Götschi, T., Garrard, J., Giles-Corti, B., 2016. Cycling as a Part of Daily Life: A Review of Health Perspectives. Transp. Rev. 36, 45–71. doi:10.1080/01441647.2015.1057877

Greenland, S., 1977. Response and follow-up bias in cohort studies. Am. J. Epidemiol. 106, 184–187. doi:10.1093/oxfordjournals.aje.a112451

Hosford, K., Fuller, D., Lear, S.A., Teschke, K., Gauvin, L., Brauer, M., Winters, M., 2018. Evaluation of the impact of a public bicycle share program on population bicycling in Vancouver, BC. Prev. Med. Reports 12, 176–181. doi:10.1016/j.pmedr.2018.09.014

Kerr, J., Emond, J.A., Badland, H., Reis, R., Sarmiento, O., Carlson, J., Sallis, J.F., Cerin, E., Cain, K., Conway, T., Schofield, G., Macfarlane, D.J., Christiansen, L.B., Van Dyck, D., Davey, R., Aguinaga-Ontoso, I., Salvo, D., Sugiyama, T., Owen, N., Mitáš, J., Natarajan, L., 2016. Perceived Neighborhood Environmental Attributes Associated with Walking and Cycling for Transport among Adult Residents of 17 Cities in 12 Countries: The IPEN Study. Environ. Health Perspect. 124, 290–298. doi:10.1289/ehp.1409466

Kristman, V., Manno, M., Côté, P., 2004. Loss to follow-up in cohort studies: how much is too much? Eur. J. Epidemiol. 19, 751–60. doi:10.1023/B

Krizek, K.J., Handy, S.L., Forsyth, A., 2009. Explaining changes in walking and bicycling behavior: challenges for transportation research. Environ. Plan. B Plan. Des. 36, 725–740. doi:10.1068/B34023

Laeremans, M., Dons, E., Avila-Palencia, I., Carrasco-Turigas, G., Orjuela, J.P., Anaya, E., Brand, C., Cole-Hunter, T., De Nazelle, A., Götschi, T., Kahlmeier, S., Nieuwenhuijsen, M., Standaert, A., De Boever, P., Int Panis, L., 2017. Physical activity and sedentary behaviour in daily life: A comparative analysis of the Global Physical Activity Questionnaire (GPAQ) and the SenseWear armband. PLoS One 12, 1–15. doi:10.1371/journal.pone.0177765

Lash, T.L., Fink, A.K., Fox, M.P., 2009. Selection Bias, in: Lash, T.L., Fox, M.P., Fink, A.K. (Eds.), Applying Quantitative Bias Analysis to Epidemiologic Data. Springer New York, New York, NY, pp. 43–57. doi:10.1007/978-0-387-87959-8_4

Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L., Weather, R.D., 2005. Cycling and the built environment, a US perspective. Transp. Res. Part D Transp. Environ. 10, 245–261. doi:10.1016/j.trd.2005.04.001

Panter, J., Costa, S., Dalton, A., Jones, A., Ogilvie, D., 2014. Development of methods to objectively identify time spent using active and motorised modes of travel to work: How do self-reported measures compare? Int. J. Behav. Nutr. Phys. Act. 11, 1–15.

Prelipcean, A.C., Susilo, Y.O., Gidófalvi, G., 2017. Collecting travel diaries : Current state of the art , best practices , and future research directions. 11th Int. Conf. Transp. Surv. Methods Collect.

Prince, S.A., Adamo, K.B., Hamel, M., Hardt, J., Connor Gorber, S., Tremblay, M., 2008. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Act. 5, 56. doi:10.1186/1479-5868-5-56

Raser, E., Gaupp-Berghausen, M., Dons, E., Anaya-Boig, E., Avila-Palencia, I., Brand, C., Castro, A., Clark, A., Eriksson, U., Götschi, T., Int Panis, L., Kahlmeier, S., Laeremans, M., Mueller, N., Nieuwenhuijsen, M., Orjuela, J.P., Rojas-Rueda, D., Standaert, A., Stigell, E., Gerike, R., 2018. European cyclists' travel behavior: Differences and similarities between seven European (PASTA) cities. J. Transp. Heal. 0–1. doi:10.1016/j.jth.2018.02.006

Sallis, J.F., Saelens, B.E., 2000. Assessment of physical activity by self-report: Status, limitations, and future directions. Res. Q. Exerc. Sport 71, 1–14. doi:10.1080/02701367.2000.11082780

Sylvia, L.G., 2015. A practical guide to measuring physical activity 114, 199–208. doi:10.1016/j.jand.2013.09.018.A

Tin Tin, S., Woodward, A., Ameratunga, S., 2014. Estimating bias from loss to follow-up in a prospective cohort study of bicycle crash injuries. Inj. Prev. 20, 322–9. doi:10.1136/injuryprev-2013-040997

Tin Tin, S., Woodward, A., Ameratunga, S., 2013. Incidence, risk, and protective factors of bicycle crashes: Findings from a prospective cohort study in New Zealand. Prev. Med. (Baltim). 57, 152–161. doi:10.1016/j.ypmed.2013.05.001

Tin Tin, S., Woodward, A., Robinson, E., Ameratunga, S., 2012. Temporal, seasonal and weather effects on cycle volume: An ecological study. Environ. Heal. A Glob. Access Sci. Source 11, 1–9. doi:10.1186/1476-069X-11-12

Vanparijs, J., Int Panis, L., Meeusen, R., de Geus, B., 2015. Exposure measurement in bicycle safety analysis: A review of the literature. Accid. Anal. Prev. 84, 9–19. doi:10.1016/j.aap.2015.08.007

Winters, M., Friesen, M.C., Koehoorn, M., Teschke, K., 2007. Utilitarian bicycling: a multilevel analysis of climate and personal influences. Am. J. Prev. Med. 32, 52–8. doi:10.1016/j.amepre.2006.08.027

Yang, Y., Diez Roux, A. V., Bingham, C.R., 2011. Variability and seasonality of active transportation in USA: evidence from the 2001 NHTS. Int. J. Behav. Nutr. Phys. Act. 8, 96. doi:10.1186/1479-5868-8-96

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., Ebooks Corporation., 2009. Mixed Effects Models and Extensions in Ecology with R, Statistics for Biology and Health. doi:10.1007/978-0-387-87458-6