

## Is active travel part of a healthy lifestyle? Results from a latent class analysis

Kroesen, Maarten

**DOI**

[10.1016/j.jth.2018.11.006](https://doi.org/10.1016/j.jth.2018.11.006)

**Publication date**

2019

**Document Version**

Accepted author manuscript

**Published in**

Journal of Transport and Health

**Citation (APA)**

Kroesen, M. (2019). Is active travel part of a healthy lifestyle? Results from a latent class analysis. *Journal of Transport and Health*, 12, 42-49. <https://doi.org/10.1016/j.jth.2018.11.006>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# 1 **Is active travel part of a healthy lifestyle? Results from a latent class analysis**

2  
3 Maarten Kroesen  
4 Delft University of Technology  
5 Faculty of Technology, Policy and Management  
6 P.O. Box 5015, 2600 GA Delft, The Netherlands  
7 Tel: +31152787183  
8 E-mail: [m.kroesen@tudelft.nl](mailto:m.kroesen@tudelft.nl)  
9

## 10 **Introduction**

11  
12 Behavioral health risk factors are major causes of morbidity and mortality worldwide. The  
13 four main risk factors, the so-called SNAP-factors, relate to smoking, nutrition, alcohol  
14 consumption and physical inactivity. A consistent finding in health research is that these  
15 behaviors tend to cluster together, thereby resulting in patterns of healthy lifestyles and  
16 unhealthy ones. In research to date, physical (in)activity is typically included using broad  
17 categories relating to the total amount of physical (in)activity. As such, it is unknown to what  
18 extent active travel behaviors (i.e. walking and cycling) as specific forms of physical activity  
19 are related to the health lifestyles.  
20

## 21 **Methods**

22  
23 In this study this knowledge gap is addressed by performing a latent class analysis based on  
24 indicators related to active travel as well as the four SNAP-factors. Data are obtained from the  
25 LISS (Longitudinal Internet Studies for the Social sciences) panel, which is based on a true  
26 probability sample of Dutch households. In total, 2,050 participants are considered in the  
27 analysis.  
28

## 29 **Results**

30  
31 Five health lifestyles are revealed and labeled as follows: consistent healthy, active  
32 commuters, physically inactive, unhealthy eaters and consistent unhealthy. The results  
33 indicate that active travel (or lack thereof) indeed forms an integral part of the consistent  
34 healthy (and unhealthy) lifestyles. In addition, lifestyle membership is found to be  
35 significantly dependent on gender, age and level of education.  
36

## 37 **Conclusion**

38  
39 For most people (70%) active travel (or lack thereof) indeed forms an integral part of these  
40 consistent healthy and unhealthy lifestyles.  
41

42 **Key words:** active travel; walking; cycling; health behaviors; SNAP factors; latent class  
43 analysis  
44

## 45 **1. Introduction**

46  
47 Behavioral health risk factors are major causes of morbidity and mortality worldwide (Lim et  
48 al., 2012). The four main behavioral risk factors are smoking, poor nutrition (lack of fruit and  
49 vegetable intake), excess alcohol consumption and physical inactivity, the so-called SNAP  
50 factors. A consistent finding in health research is that these behaviors tend to co-occur or

51 cluster together, thereby resulting in patterns of healthy lifestyles and unhealthy ones (Noble  
52 et al., 2015; McAloney, 2013; Meader et al., 2016). This empirical finding suggests that an  
53 approach is needed which focuses on tackling multiple risk factors simultaneously, instead of  
54 strategies that focus on changing isolated behaviors (Prochaska et al., 2008). Such an  
55 approach is further supported by evidence indicating that the risk factors have synergistic effects  
56 on health, i.e. combinations of risk behaviors are more detrimental to health than their  
57 individual effects (French et al., 2008; Poortinga, 2007).

58 To support the development of a comprehensive strategy it is necessary to know which  
59 risk factors indeed cluster together in particular contexts and populations. Indeed, already a  
60 fair amount of research has been devoted to this subject (Noble et al., 2015; McAloney, 2013;  
61 Meader, 2016). Based on review of 56 studies, Noble et al. (2015) found that the majority of  
62 studies (81%) reported a relatively 'healthy' cluster, which is characterized by the absence of  
63 any behavioral risk factors. In addition, half of the studies revealed the presence of a  
64 consistent unhealthy lifestyle, in which all risk factors (smoking, poor nutrition, excess  
65 alcohol consumption and high physical inactivity) were prevalent (Noble et al., 2015).

66 To date, physical (in)activity is typically included using broad categories relating to  
67 the amount of time spent sedentary and/or the amount of general physical activity. As such, it  
68 is unknown to what extent active travel behaviors (i.e. walking and cycling) as specific forms  
69 of physical activity are related to the different lifestyles. This study aims to address this  
70 knowledge gap and assess the extent active travel (or the lack thereof) is part of general  
71 healthy (or unhealthy) lifestyles. If active travel is indeed part of a comprehensive healthy  
72 lifestyle pattern(s), this has important policy implications, as it suggests that active travel may  
73 be increased by stimulating the more generic health lifestyles.

74 To reveal the health lifestyles, a latent class analysis is performed based on indicators  
75 related to active travel as well as a general measure of physical inactivity and indicators  
76 related to the three other behavioral risk factors (smoking, alcohol and nutrition). Socio-  
77 demographic variables (gender, age and education level) and the Body-Mass Index (BMI) are  
78 included as covariates to further profile the classes. The data used for the analysis are  
79 obtained from two surveys (related to health and travel behavior) conducted in the LISS  
80 (Longitudinal Internet Studies for the Social sciences) panel<sup>1</sup>. This panel is based on a true  
81 probability sample of Dutch households. 2,050 individuals completed both surveys and are  
82 included in the analysis.

83 In the following, relevant empirical findings and theoretical mechanisms will be  
84 discussed. After that, the empirical study will be introduced and its results discussed. In the  
85 final section, the conclusions are summarized and several policy recommendations are  
86 formulated.

87

## 88 **2. Empirical and theoretical background**

89

90 Active travel (walking and cycling) is increasingly being recognised as a potentially effective  
91 means of increasing physical activity levels and thereby contribute to physical and mental  
92 health (Sallis, 2004; Frank et al., 2006; Pucher et al., 2010). Active travel can often easily be  
93 incorporated in the daily routine. In addition, there is much scope for active travel to help  
94 people meet recommended physical activity levels. In the US, for example, 27% of all trips in  
95 2009 were shorter than 1 mile, but only 36% of those short trips were made by walking or  
96 cycling (Buehler et al., 2011).

97 Research related to active travel is largely driven by two questions: (1) what are the  
98 health effects of active travel? and (2) what are the causes of active travel? Multiple

---

<sup>1</sup> [www.lissdata.nl](http://www.lissdata.nl)

99 disciplines are involved in answering these two questions and the resulting literatures are vast.  
100 Relevant potential outcomes include increased total physical activity, reduced obesity,  
101 increased fitness and increased psychological well-being (see Oja et al. (2011), Wanner et al.  
102 (2012) and Saunders et al. (2013) for relevant reviews). Research related to the determinants  
103 of active travel has focused on the role of the built environment (e.g. residential density) and  
104 available bicycle and pedestrian infrastructure (a review of reviews is provided by Ding &  
105 Gebel (2012)). Also psychological factors (perceived environmental characteristics or  
106 attitudes and preference) have been considered, albeit to a lesser extent (Panter & Jones,  
107 2010; Heinen et al., 2011).

108 Given that active travel is important health behaviour, it is relevant to know whether  
109 and to what extent it clusters with other health behaviours. As discussed above, the literature  
110 regarding the clustering of health behaviors is already quite extensive. In the health domain,  
111 many studies adopt an empirical perspective and focus on revealing the existing behavioral  
112 patterns. In this regard, consistent healthy and unhealthy clusters have been reported (Noble et  
113 al., 2015; McAloney, 2013; Meader, 2016). Moreover, it has been observed that some  
114 behaviors are more strongly correlated than others. For example, excessive alcohol use and  
115 smoking are typically found to cluster together, and the same holds for physical inactivity and  
116 poor diet (Meader et al., 2016). Other combinations are less common. For example, no  
117 association is generally found between excessive alcohol intake and physical inactivity (Noble  
118 et al., 2015).

119 Research has also focused on the determinants of profile membership. Here, the most  
120 consistent finding is that lower socio-economic status in terms of education level, income or  
121 type of occupation is associated with membership of more risky clusters. Regarding other  
122 socio-demographic characteristics the evidence is mixed. Males tend to have a higher  
123 probability than females of being a member of an unhealthy cluster, but in general gender is  
124 found to be a weak predictor of cluster membership. For age, the effects are also inconsistent.  
125 Although some studies report that younger age is associated with multiple risk behaviors (see  
126 e.g. Poortinga, 2007), most studies report non-significant findings (Noble et al., 2015) and  
127 some even indicate that older age is associated with riskier clusters (see e.g. Lee et al., 2012).

128 The theoretical mechanisms which supposedly underlie the clustering of health  
129 behaviors are still poorly understood. In the literature at least three categories of theoretical  
130 mechanisms have been identified, namely biological, psychological and sociological.  
131 Examples of biological mechanisms include notions that smoking (nicotine) may counteract  
132 the depressant effects of alcohol use or that heavy smoking may reduce lung function and  
133 thereby discourage physical activity. These biological explanations can account for the  
134 observation that some health behaviors are more strongly correlated than others.

135 Psychological mechanisms generally relate to the idea that health behaviors (at least  
136 partially) are the outcomes of a rational choice process, which, for example, is assumed by  
137 psychological theories such as the Health Belief Model (Janz and Becker, 1984) or the Theory  
138 of Planned Behavior (Ajzen, 1991). Clustering would be expected if individuals, via such a  
139 rational process, consistently reach the same (bad/good) decisions regarding whether to adopt  
140 certain health behaviors or not.

141 Finally, sociological studies draw attention to broader institutional/structural  
142 explanations. An example is the health lifestyle theory developed by Cockerham (2005). This  
143 theory postulates that particular health lifestyles originate from the interplay of life choices  
144 (agency) and life chances (structure). The life chances are, amongst others, determined by  
145 class circumstances and living conditions, which may operate as either constraints or enablers  
146 of certain health lifestyles, resulting in consistent healthy or unhealthy behavioral patterns.  
147 Hence, sociological theories (like the health lifestyle theory) emphasize the fact that some  
148 groups in society have fewer life chances than others (due to a lack of resources, time, or

149 access to healthy food and exercise opportunities) and can therefore be expected to engage in  
150 multiple health risk behaviors.

151 While empirical research so far has shed little light as to which theory (or combination  
152 of theories) actually explains the clustering of health behaviors, it concluded that various  
153 mechanisms can be identified that may be responsible. Yet, in future work, it would be  
154 interesting to try and uncover which mechanisms are actually most relevant. This issue will be  
155 returned to in the discussion.

156 Both psychological and sociological theories ascribe an important role to the level of  
157 education as a determinant of engaging in multiple health behaviors. Within psychological  
158 models, it may be expected that higher educated people have more knowledge of health  
159 behaviors (and can also more easily acquire new knowledge), which arguably will result in  
160 better (healthier) decisions. In sociological models, the level of education is an important  
161 indicator of class (in postmodern societies) which, as discussed above, may enable or  
162 constrain certain health behaviors. For example, higher educated people may engage in  
163 healthy lifestyle, because it is perceived as the class norm, thereby also allowing them to  
164 distinguish themselves from other ('lower') classes.

165 This study contributes to the health literature by empirically assessing to what extent  
166 active travel (or lack thereof), as a specific form of physical (in)activity, is part of (un)healthy  
167 lifestyles. Typically, broad measures of physical activity are considered, relating to the total  
168 amount of sedentary time or general physical activity. As such, it is as of yet unknown to what  
169 extent forms of active travel are related to health lifestyles.

170 This study also aims to contribute to the transportation literature. In transportation  
171 research, active travel is often (implicitly) conceptualized as an 'environmentally friendly'  
172 behavior, as opposed to a 'health-enhancing' behavior. As such, research typically considers  
173 psychological constructs related to environmental attitudes or beliefs to explain active travel  
174 (see e.g. Bamberg and Möser, 2007). While correlations between such attitudes and active  
175 travel are typically found, it might be that (some) individuals also engage in active travel as a  
176 means to stay fit and healthy.

177 One way to test this idea is to measure both of these psychological motivations and  
178 assess their effects on active travel. Empirically, however, as shown by the research of Heinen  
179 et al. (2011), motivations related to environmental and health benefits are highly  
180 intercorrelated (both motivations were actually found to load on the same factor). Hence, it  
181 has proven difficult to assess the (unique) contribution of health motivations in explaining  
182 variation in active travel. In addition, the motivations may also be adopted post-hoc, as a way  
183 to justify the behavior (Kroesen et al., 2017). Hence, the direction of causation always  
184 remains uncertain.

185 By assessing the extent to which active travel is part of healthy lifestyles this problem  
186 can be addressed to some extent. More specifically, should active travel (indeed) occur mainly  
187 within comprehensive healthy lifestyles, this would suggest that health motivations indeed  
188 play a relevant role (in addition to the possible role of environmental motivations). While  
189 such evidence does not prove that health motivations do play a role, any lack of clustering of  
190 active travel and other health behaviors would quite definitively prove that health motivations  
191 do not play a role. So the present study should be regarded as another relevant piece of the  
192 puzzle with regard to the role of health motivations in determining level of active travel.

193

### 194 **3. Method**

195

#### 196 **3.1 Data and measures**

197

198 The data used for the analysis are drawn the LISS (Longitudinal Internet Studies for the  
 199 Social sciences) panel, which is based on a true probability sample of Dutch households.<sup>2</sup>  
 200 From this panel data from two surveys are combined, one survey on travel behavior  
 201 conducted in July 2013<sup>3</sup> and one on health conducted in November 2013<sup>4</sup>. For the travel  
 202 behavior survey 2,980 panel members were invited and 2,370 responded (response rate  
 203 79.5%) and for the health survey 6,217 were invited and 5,379 responded (response rate  
 204 79.5%). In total, 2,050 individuals completed both of these surveys and are included in the  
 205 analysis.

206 Since (additional) selection bias may have been introduced by considering only people  
 207 who participated in both surveys (although both individually have quite high response rates),  
 208 several representativeness tests were conducted. Table 1 presents the sample distributions of  
 209 three sociodemographic variables (gender, age and level of education) and Body-Mass Index  
 210 (BMI) and the respective population distributions retrieved from Statistics Netherlands. The  
 211 results indicate that the sample is representative for the population of Dutch adults with  
 212 respect to gender and BMI (no significant differences). However, the mean age in the sample  
 213 is 3.9 years higher than in the population, and also the level of education is (on average)  
 214 higher in the sample compared to the population. The implications of these results with  
 215 respect to the findings will be reflected upon in the concluding section.  
 216

217 **Table 1. Sample and population distributions of sociodemographic variables and BMI**  
 218

		Sample	Population <sup>a</sup>	Test of significance
Gender (%)	Male	47	49	$\chi^2=2.3$ , $df=1$ , $p$ -value=0.127
	Female	53	51	
Age	Mean	51.6	47.7	$t=10.5$ , $df=2049$ , $p$ -value=0.000
Level of education (%)	Low	32	33	$\chi^2=3.5$ , $df=2$ , $p$ -value=0.000
	Intermediate	36	40	
	High	33	27	
BMI (kg/m <sup>2</sup> ) (%)	Underweight (<18.5)	2	2	$\chi^2=6.5$ , $df=3$ , $p$ -value=0.090
	Normal weight (18.5-25)	48	50	
	Overweight (25-30)	37	36	
	Obese (>30)	13	12	

219 <sup>a</sup> Data retrieved from Statistics Netherlands (<http://statline.cbs.nl/Statweb/>)  
 220

221 Active travel was operationalized using three indicators: the distance travelled by  
 222 bicycle in a regular week (measured on a 5-point ordinal scale), the number of days that the  
 223 respondent spent at least 10 minutes walking in the past week and a dummy indicating  
 224 whether the respondent is an active commuter or not (walking or cycling to work or school).

225 The SNAP-factors were measured as follows. Smoking was operationalized using a  
 226 simple indicator whether the respondent currently uses tobacco (smoking cigarettes, cigars or  
 227 pipes) or not. Excessive alcohol consumption was operationalized with the following  
 228 question: how often did you have a drink containing alcohol over the last 12 months?  
 229 Respondents could indicate their answer on an 8-point ordinal scale ranging from (1) not at all  
 230 over the last 12 months to (8) almost every day. Only those falling in the highest category (8)  
 231 were considered as excessive drinkers. Hence, the original scale was recoded into a binary  
 232 variable indicating excessive consumption or not. The nutrition factor was operationalized  
 233 using two questions relating to fruit and vegetable consumption. For these questions 6-point  
 234 ordinal scales were used ranging from (1) never to (6) every day. Here, the last two  
 235 categories, (5) 5 to 6 times per week and (6) every day, were considered as indicative of a

<sup>2</sup> Details on the panel can be found at [www.lissdata.nl](http://www.lissdata.nl)

<sup>3</sup> See [https://www.dataarchive.lissdata.nl/study\\_units/view/584](https://www.dataarchive.lissdata.nl/study_units/view/584)

<sup>4</sup> See [https://www.dataarchive.lissdata.nl/study\\_units/view/509](https://www.dataarchive.lissdata.nl/study_units/view/509)

236 good diet. Hence, these two variables were also recoded into binary variables. Finally, in  
237 addition to the indicators related to active travel, a broad measure of physical inactivity was  
238 included, namely the number of hours spend on sedentary activities on a regular day. For this  
239 question a 4-point ordinal scale was used ranging from (1) 0-3 hours per day to (4) 10 or more  
240 hours per day. In total, eight indicators were used in the analysis.

241 Note that the used indicators for active travel capture active travel for both  
242 transportation and leisure purposes. Given that general physical activity is operationalized as  
243 the time spend sedentary (as opposed to some form of physical activity), there is no overlap  
244 with this indicator.

245 Unfortunately, the used indicators were not operationalized in such a way that general  
246 recommended thresholds for healthy (unhealthy) behaviors could be adopted. For example,  
247 WHO's recommendation regarding vegetable/fruit consumption consists of eating 400 grams  
248 of fruits and vegetables per day. In a similar fashion, regarding physical activity, WHO  
249 recommends that adults aged 18-64 should do at least 150 minutes of moderate-intensity  
250 aerobic physical activity throughout the week. Since the scales based upon which these  
251 thresholds are defined do not align with the answering scales used in the survey, the  
252 behavioral patterns (resulting from the analysis) cannot be interpreted in an absolute sense  
253 (i.e. based on the accepted guidelines), as being either 'healthy' or 'unhealthy'. Instead, the  
254 resulting patterns should be interpret in a relatively sense, as being relatively more or less  
255 health than other patterns and/or the sample average.

256 Next to the indicators, the socio-demographic characteristics and BMI were included  
257 as covariates in the model. To assess the significance of these covariates the 3-step procedure  
258 was applied (Vermunt, 2010). An advantage of this procedure over the 1-step approach  
259 (where covariates are directed included to predict class membership) is that the covariates will  
260 not interfere with the measurement part of the model, i.e. classification is solely based on the  
261 indicators (and not the covariates). The 3-step procedure basically consists of the following  
262 steps: (1) estimation of the model based on the indicators only, (2) probabilistic assignment of  
263 subjects to latent classes (the posterior membership probabilities) and (3) estimation of the  
264 effects of the covariates on latent class membership, corrected for the classification error to  
265 prevent bias. The procedure allows the researcher to establish the effects (and significance) of  
266 the covariates (corrected for measurement errors), while not letting the covariates interfere  
267 with the classification based on the indicators (Vermunt, 2010).

### 268 **3.2 Model estimation**

269

270 Latent class modelling has several advantages over traditional clustering techniques, such as  
271 K-means cluster analysis (Magidson and Vermunt, 2002). One particular advantage is that  
272 nominal and ordinal indicators can be used (in addition to continuous outcomes), which were  
273 also present in the current application. Within the model specification all indicators were  
274 specified as either nominal or ordinal. Latent Gold 5.1 was to estimate the latent class models  
275 (Vermunt and Magidson, 2013).

276 The goal of the latent class analysis is to find the most parsimonious model, i.e. with  
277 the smallest number of latent classes, which adequately describes the associations between the  
278 indicators. To identify the optimal model, subsequent models were estimated with 1 through 8  
279 latent classes. Table 2 presents the fit of these models in terms of the Bayesian information  
280 criterion (BIC), a statistic which weighs model fit and model parsimony, and the sum of the  
281 bivariate residuals (BVRs), indicating the total amount of association remaining between the  
282 indicators after accounting for the latent class variable (Vermunt and Magidson, 2013).

283 Based on the BIC statistic (being lowest in the 3-class model) the 3-class model  
284 should be considered optimal. However, in this solution, significant bivariate residuals  
285 (>3.84) remained between the indicators (note that since the bivariate residuals are chi-

286 squared distributed with one degree of freedom, a value of 3.84 corresponds to the critical chi-  
287 square value at the 5% level of significance (Vermunt and Magidson, 2013). These bivariate  
288 residuals were not reduced up until the 5-class solution. Since the 5-class solution also  
289 provided additional relevant substantive insights over the 3-class model, the decision was  
290 made to consider this solution as optimal. In the next section this solution will therefore be  
291 interpreted substantively.

292

293 **Table 2. Model fit of the latent class models**

294

No. of classes	Npar	LL	BIC(LL)	Sum of BVRs
1	19	-15562.6	31270.0	492.4
2	28	-15447.4	31108.3	261.9
3	37	-15368.7	31019.6	85.7
4	46	-15343.3	31037.3	63.8
5	55	-15318.8	31057.0	26.1
6	64	-15302.3	31092.7	25.5
7	73	-15289.1	31134.8	13.8
8	82	-15273.8	31172.9	7.6

295

296 Npar = number of model parameters

297 LL = final log-likelihood

298 BIC(LL) = Bayesian information criterion (based on log-likelihood).

299 Total BVR = sum of the bivariate residuals

300

301



302 **4. Results**

303

304 Table 3 presents the class sizes and the profiles of the five classes. To aid the  
 305 interpretation the final column presents the sample distributions. Overall, the five classes  
 306 represented well-interpretable patterns. Note that, although the estimation of the measurement  
 307 model (model with only indicators) and structural model (model with covariates) occurs  
 308 consecutively (in line with the 3-step procedure), the results of the third step are already  
 309 included here. Hence, the distributions and tests of significance of the covariates are included  
 310 in the profile output (at the bottom).

311

312 **Table 3. Profiles of the 5 latent classes and the sample distributions**

313

		1	2	3	4	5	Sample
<b>Cluster Size (%) N=2,050</b>		41	19	18	11	10	
<b>Indicators</b>							
Distance (in kilometer) travelled by bicycle in a regular week (%)	0	6	3	37	20	34	16
	1-10	24	14	46	43	46	30
	11-20	19	15	11	18	12	16
	21-40	23	25	4	11	5	17
	>40	28	44	2	8	2	21
No. of days with more than 10 minutes walking in past week	Mean	3.4	3.9	1.9	2.4	1.9	3.0
Active commuter (%)	No	75	28	100	72	92	72
	Yes	25	72	0	28	8	28
No. of hours sedentary on a regular day (%)	0-3	21	11	9	23	18	17
	4-6	43	34	31	43	41	39
	7-9	20	25	26	20	22	22
	10 or more	16	29	34	14	19	22
Currently smokes tobacco (%)	No	95	83	95	90	4	83
	Yes	5	17	5	10	96	17
Excessive drinking (drinking almost every day in past year) (%)	No	84	90	83	97	70	85
	Yes	16	10	17	3	30	15
Frequency of eating vegetables (5 times per week or more) (%)	No	23	34	26	95	43	36
	Yes	77	66	74	5	57	64
Frequency of eating fruit (5 times per week or more) (%)	No	15	57	36	86	62	40
	Yes	85	43	64	14	38	60
<b>Covariates</b>							
Gender (%) (Wald=28.4, p=0.00)	Male	40	50	55	56	48	47
	Female	60	50	45	44	52	53
Age (Wald 121.0, p=0.00)	Mean	61.9	39.5	49.3	42.5	50.6	51.6
Level of education (%) (Wald 56.9, p=0.00)	Low	38	20	24	40	38	32
	Intermediate	29	44	32	44	40	36
	High	33	36	43	17	22	33
BMI (kg/m <sup>2</sup> ) (%) (Wald= 21.5, p=0.04)	Underweight (<18.5)	1	5	0	1	2	2
	Normal weight (18.5-25)	45	60	41	45	51	48
	Overweight (25-30)	43	25	37	40	34	37
	Obese (>30)	11	10	22	14	14	13

314

315 Note: some column values may not add up to 100% due to rounding

316

317 The first two classes represent relatively healthy lifestyles. Subjects belonging to the first  
 318 class (41% of the sample) cycle and walk above the sample average. A substantial portion of  
 319 the subjects is also an active commuter (25%). Compared to the other classes, the time spend  
 320 sedentary (on average per day) is among the lowest in this class. Relatively few engage in  
 321 smoking (5%) and drinking (16%) and the frequencies of eating vegetables and fruit (5 times  
 322 or more per week) are high (77% and 85% respectively). Overall, the first class represents a  
 323 relatively *consistent healthy* lifestyle.

324 While the second class (19% of the sample) also represents a relatively healthy  
325 lifestyle, there are several distinct differences with the first class. Firstly, compared to the first  
326 class, levels of cycling and walking are higher in this class. In addition, 75% of the subjects  
327 are *active commuters*. At the same time, the amount of sedentary time is also higher. Thirdly,  
328 compared to the first class, smoking occurs relatively more frequently (17%), while drinking  
329 occurs less frequently (10%). Overall, however, these levels are still at the low end of the  
330 spectrum. A similar pattern occurs with respect to vegetable and fruit intake, which is still  
331 quite high, but again slightly lower than in the first class.

332 The third class (18% of the sample) scores especially poor on the indicators of  
333 *physical inactivity*. Compared to the other classes, the levels of cycling and walking are  
334 lowest, while the time spent sedentary is the highest. The levels of smoking, drinking and  
335 vegetable/fruit intake are, however, comparable to the first (healthy) class. Hence, only in  
336 terms of physical inactivity does this pattern represent an unhealthy lifestyle.

337 The fourth class (11% of the sample) has low levels of cycling and walking, but still a  
338 substantial portion of active commuters (28%). In addition, the amount of sedentary time is  
339 relatively low and smoking/drinking also occurs relatively infrequently. The distinct feature  
340 of this class is the low level of vegetable and fruit intake; 95% and 86% respectively does not  
341 meet the threshold of eating vegetables/fruits 5 times or more per week. Hence, people with  
342 this pattern can be described as *unhealthy eaters*.

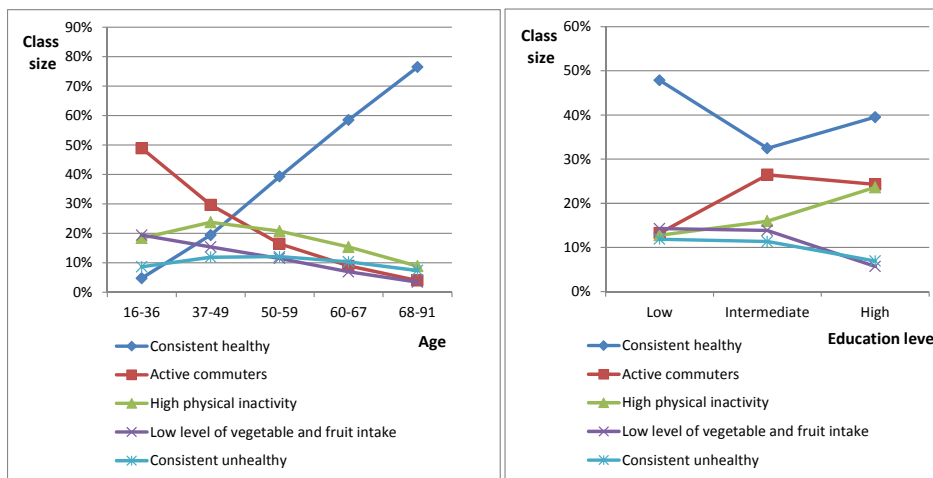
343 Finally, the last class (10% of the sample) represents a relatively *consistent unhealthy*  
344 lifestyle. Subjects in this class engage little in active travel. In addition, smoking and drinking  
345 levels are high (96% and 30% respectively), while the levels of fruit and vegetable intake are  
346 low.

347 All four covariates were found to be significant ( $p < 0.05$ ). The results indicate that  
348 females are relatively more likely to belong to the first ('consistent healthy') lifestyle, while  
349 males are more likely to belong to the fourth ('low level of vegetable and fruit intake') class.  
350 Subjects in the first class ('consistent healthy') are on average relatively old (mean=61.9),  
351 while subjects of the second ('active commuters') and fourth ('low level of vegetable and fruit  
352 intake') class are relatively young (mean~40). Probably, the lack of active commuting in the  
353 first 'consistent healthy' class is (partly) due to the high average age in the class, resulting in  
354 the fact that relatively many in the class do not have to commute for work. Overall, the  
355 education level is higher in the healthier classes (1 and 2) compared to the unhealthier ones (4  
356 and 5). Interesting, however, the education level is highest in the third class ('high physical  
357 inactivity'), which is probably due to the fact that people in this class are more likely to have  
358 an (sedentary) office job. In line with the level of physical activity, the BMI is highest in the  
359 third class (22% obese), followed by the relatively unhealthy classes (class 4 and 5, ~14%  
360 obese) and the healthier ones (class 1 and 2, ~10% obese).

361 A more intuitive way to interpret the effects of the covariates, which is also more in  
362 line with the underlying conceptualization that the covariates influence class membership, is  
363 to calculate and assess the predicted class membership probabilities for various levels of the  
364 covariates (while holding the other covariates at their mean value). This is done for age and  
365 level of education, which have the strongest effects on class membership. Figure 1 plots class  
366 membership as a function of these variables.

367 With respect to age it becomes clear that, over the life course, membership of the  
368 consistent healthy lifestyle (class 1) steadily increases. This growth occurs mainly at the  
369 expense of the active commuter lifestyle (class 2), but also of the other three lifestyles. Hence,  
370 with increasing age people mainly transition from the active commuter lifestyle to the  
371 consistent healthy lifestyle, thereby reducing their active travel but generally improving their  
372 health behaviors in terms of (not) smoking and their diet. This reduction in active travel is  
373 plausible given the increasing difficulty of traveling actively at older age.

374 The effects of level of education are somewhat inconsistent with previous research.  
 375 Typically, higher education is found to be associated with healthier lifestyles (Noble et al.,  
 376 2015). Indeed, the findings here also show that the probability of being a member of the  
 377 consistent unhealthy lifestyle is lowest (7%) when the level of education is highest. Yet, as  
 378 discussed above, membership of the relatively unhealthy ‘physical inactive’ profile also  
 379 increases consistently with the level of education. Another interesting finding is that the level  
 380 of education strongly influences the class membership distributions of the first two classes,  
 381 i.e. the consistent healthy and the active commuter lifestyle. Especially in moving from the  
 382 low to the intermediate level of education a large shift in these classes can be observed,  
 383 whereby the active commuter pattern strongly increases at the expense of the consistent  
 384 healthy lifestyle. Hence, in the Dutch context, active commuting is strongly linked to the level  
 385 of education.  
 386



387  
 388  
 389 *Figure 1. Predicted class membership probabilities conditional on age (left) and education*  
 390 *level (right)*  
 391

## 392 5. Discussion

393  
 394 Taken together the results of the research are in line with previous findings, although some  
 395 particular unexpected results are revealed as well. Firstly, in line with previous studies, at both  
 396 ends of the spectrum evidence of clustering is found. Around 60% of the sample (class 1 and  
 397 2) has a consistent healthy lifestyle, while 10% has a consistent unhealthy lifestyle (class 5).  
 398 In addition, in line with the main objective of the present study, it is shown that for most  
 399 people (70%) active travel (or lack thereof) indeed forms an integral part of these consistent  
 400 healthy and unhealthy lifestyles, i.e. with high engagement in the overall healthy lifestyles  
 401 (class 1 and 2) and low engagement in the overall unhealthy lifestyle (class 5). Yet, two  
 402 classes (3 and 4) are revealed for which this does not hold. Especially the third class stands  
 403 out in this regard, as it represents an overall healthy lifestyle with the exception of physical  
 404 activity (including active travel).

405 The effects of the covariates are also in line with findings of previous studies, i.e. on a  
 406 whole, the probability of being a member of one of the healthier classes increases with being  
 407 female, level of education and age. Still, here as well, some unexpected findings are revealed.  
 408 For example, the finding that the level of education is highest (on average) in the relatively  
 409 unhealthy ‘physical inactivity’ profile (class 3). Another peculiar finding is that age does not  
 410 strongly affect the probability of being a member of this class and an also the effect of the  
 411 level of education is relatively weak. This somewhat contrasts previous research which

412 reported stronger relationships between health lifestyles and (especially) the level of  
413 education. It may be speculated that the level of education in general is quite high in the  
414 Netherlands and in this sample in particular (as the representativeness analysis has shown),  
415 thereby reducing variation in the variable and weakening the effect on the health lifestyles.

416 Finally, some limitations of this study and related avenues for future research can be  
417 identified. The most important limitation is that the health risk behaviors are based on self-  
418 report, which, due to social desirability bias, tend to underestimate the prevalence of health  
419 risk behaviors and overestimate healthy behaviors (Newell et al., 1999). Yet, in many cases  
420 they remain the only feasible method of collection (Noble et al., 2015). A related issue, is that  
421 the answering scales did not match those used in setting general guidelines as to what  
422 constitutes healthy behavior (or not), a point which was discussed in section 3. Given that this  
423 is the case, it is difficult to interpret the patterns in an absolute sense, i.e. as being either  
424 healthy or not. Ideally, future research should be based on more accurate and objective  
425 measurements of active travel and the SNAP factors. For example, including measurements  
426 on actual fruit and vegetable intake or on the actual time spend on active travel. Considering  
427 the measurement of active travel in particular, in the transportation studies this is often  
428 measured using travel diaries. Probably, this approach will yield more reliable estimates of  
429 active travel compared to general questions about (weekly) travel behavior by various modes.  
430 It would be interesting to use data from such studies in future health research.

431 A second limitation relates to the sample representativeness and the generalizability of  
432 the findings. The representativeness analysis has shown that older and higher educated people  
433 are overrepresented in the sample compared to the population. Since age and education level  
434 positively influence the probability of being a member of the ‘consistent healthy’ class, this  
435 class is probably overrepresented in the sample compared to the population. At the same time,  
436 it is reassuring to see that no bias exists with respect to BMI, suggesting that no health-related  
437 selection mechanisms were at work (i.e. those with poor health being more inclined to deny  
438 participation).

439 A third limitation relates to the cross-sectional nature of the data, making it impossible  
440 to draw causal inferences and/or assess intra-individual change over time. This limitation may  
441 be addressed in future work as data are available from multiple waves (years) in the LISS  
442 panel. Specifically, a latent transition model (Collins and Lanza, 2013) may be estimated to  
443 reveal how people transition between the different latent classes over time and assess which  
444 factors and/or events trigger transitions to (un)healthy lifestyles.

445 Finally, similar to most studies in the health literature which adopt an  
446 empirical/descriptive approach, i.e. focused on revealing which clusters exist in the  
447 population, this study does not provide an answer to the question why these clusters exist.  
448 While qualitative research may shed light on this question, a quantitative approach would be  
449 to additionally include various (stated) motivations for engaging in the respective health  
450 behaviors in the latent class model. It may then be assessed what are the primary motivations  
451 for (not) engaging in multiple health behaviors. Yet, at this point, one (again) runs into the  
452 trouble that motivations may be adopted post-hoc instead of driving the behaviors.

453

## 454 **References**

455

456 Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human*  
457 *decision processes*, 50(2), 179-211.

458 Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new  
459 meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of*  
460 *environmental psychology*, 27(1), 14-25.

461 Buehler, R., Pucher, J., Merom, D., & Bauman, A. (2011). Active travel in Germany and the  
462 US: contributions of daily walking and cycling to physical activity. *American Journal of*  
463 *Preventive Medicine*, 41(3), 241-250.

464 Cockerham, W. C. (2005). Health lifestyle theory and the convergence of agency and  
465 structure. *Journal of health and social behavior*, 46(1), 51-67.

466 Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With*  
467 *applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons.

468 Ding, D., & Gebel, K. (2012). Built environment, physical activity, and obesity: what have we  
469 learned from reviewing the literature? *Health & Place*, 18(1), 100-105.

470 Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W.  
471 (2006). Many pathways from land use to health: associations between neighborhood  
472 walkability and active transportation, body mass index, and air quality. *Journal of the*  
473 *American planning Association*, 72(1), 75-87.

474 French, S., Rosenberg, M., & Knuiman, M. (2008). The clustering of health risk behaviours in  
475 a Western Australian adult population. *Health Promotion Journal of Australia*, 19(3), 203-  
476 209.

477 Heinen, E., Maat, K., & Van Wee, B. (2011). The role of attitudes toward characteristics of  
478 bicycle commuting on the choice to cycle to work over various distances. *Transportation*  
479 *research part D: transport and environment*, 16(2), 102-109.

480 Janz, N. K., & Becker, M. H. (1984). The health belief model: A decade later. *Health*  
481 *education quarterly*, 11(1), 1-47.

482 Kroesen, M., Handy, S., & Chorus, C. (2017). Do attitudes cause behavior or vice versa? An  
483 alternative conceptualization of the attitude-behavior relationship in travel behavior modeling.  
484 *Transportation Research Part A: Policy and Practice*, 101, 190-202.

485 Lee, Y., Back, J. H., Kim, J., Byeon, H., Kim, S., & Ryu, M. (2012). Clustering of multiple  
486 healthy lifestyles among older Korean adults living in the community. *Geriatrics &*  
487 *gerontology international*, 12(3), 515-523.

488 Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., ... & Aryee,  
489 M. (2012). A comparative risk assessment of burden of disease and injury attributable to 67  
490 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the  
491 *Global Burden of Disease Study 2010*. *The lancet*, 380(9859), 2224-2260.

492 Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-  
493 means. *Canadian Journal of Marketing Research*, 20(1), 36-43.

494 McAloney, K., Graham, H., Law, C., & Platt, L. (2013). A scoping review of statistical  
495 approaches to the analysis of multiple health-related behaviours. *Preventive medicine*, 56(6),  
496 365-371.

497 Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., ... & Sowden,  
498 A. J. (2016). A systematic review on the clustering and co-occurrence of multiple risk  
499 behaviours. *BMC public health*, 16(1), 657.

500 Newell, S. A., Girgis, A., Sanson-Fisher, R. W., & Savolainen, N. J. (1999). The accuracy of  
501 self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in  
502 the general population: a critical review. *American journal of preventive medicine*, 17(3), 211-  
503 229.

504 Noble, N., Paul, C., Turon, H., & Oldmeadow, C. (2015). Which modifiable health risk  
505 behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol  
506 and Physical activity ('SNAP') health risk factors. *Preventive medicine*, 81, 16-41.

507 Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., & Kohlberger, T.  
508 (2011). Health benefits of cycling: a systematic review. *Scandinavian journal of medicine &*  
509 *science in sports*, 21(4), 496-509.

510 Panter, J. R., & Jones, A. (2010). Attitudes and the environment as determinants of active  
511 travel in adults: what do and don't we know? *Journal of Physical Activity and Health*, 7(4),  
512 551-561.

513 Poortinga, W. (2007). The prevalence and clustering of four major lifestyle risk factors in an  
514 English adult population. *Preventive medicine*, 44(2), 124-128.

515 Prochaska, J. J., Spring, B., & Nigg, C. R. (2008). Multiple health behavior change research:  
516 an introduction and overview. *Preventive medicine*, 46(3), 181-188.

517 Pucher, J., Buehler, R., Bassett, D. R., & Dannenberg, A. L. (2010). Walking and cycling to  
518 health: a comparative analysis of city, state, and international data. *American journal of public*  
519 *health*, 100(10), 1986-1992.

520 Sallis, J. F., Frank, L. D., Saelens, B. E., & Kraft, M. K. (2004). Active transportation and  
521 physical activity: opportunities for collaboration on transportation and public health research.  
522 *Transportation Research Part A: Policy and Practice*, 38(4), 249-268.

523 Saunders, L. E., Green, J. M., Petticrew, M. P., Steinbach, R., & Roberts, H. (2013). What are  
524 the health benefits of active travel? A systematic review of trials and cohort studies. *PLoS*  
525 *One*, 8(8), e69912.

526 Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step  
527 approaches. *Political analysis*, 18(4), 450-469.

528 Vermunt, J. K., & Magidson, J. (2013). Technical guide for Latent GOLD 5.0: Basic,  
529 advanced, and syntax. *Belmont, MA: Statistical Innovations Inc.*

530 Wanner, M., Götschi, T., Martin-Diener, E., Kahlmeier, S., & Martin, B. W. (2012). Active  
531 transport, physical activity, and body weight in adults: a systematic review. *American journal*  
532 *of preventive medicine*, 42(5), 493-502.

533