Transportation
https://doi.org/10.1007/s11116-022-10266-z

How does self-assessed health status relate to preferences for cycling infrastructure? A latent class and latent variable approach

Tomás Rossetti[1] · Ricardo Daziano[2]

Abstract
This study aims to understand how self-assessed health status relates to preferences for cycling infrastructure. An integrated latent class and latent variable choice model is fitted using responses to a stated preference experiment from a panel of New York City residents (N = 801). Estimates show that people with stated good physical health tend to have preference parameters similar to those of experienced cyclists. This result means that the provision of cycling infrastructure with the purpose of attracting non-cyclists also has the potential of attracting those with worse health outcomes. This result suggests a double benefit coming from car use reduction and lower health spending.

Keywords Transportation and health · Cycling · Latent variable · Latent class

Introduction

The past two decades have seen increasing research interest in the analysis of cyclists' preferences for cycling infrastructure (Nello-Deakin 2020; Pucher and Buehler 2008). These studies have used different methods to identify the built environment characteristics that are preferred by cyclists, and that could therefore be exploited to encourage a broader modal shift toward sustainable transportation. The vast consensus is that cyclists prefer infrastructure that is separated from traffic, as well as shorter and more direct routes (Buehler and Dill 2016).

Even though this consensus may be true for the population as a whole, there are significant differences both within cyclists and non-cyclists that should be considered during policy formulation. For example, a review carried out by Aldred et al. (2016) shows that women and the elderly tend to have a stronger preference for segregated cycling

✉ Tomás Rossetti
ter58@cornell.edu

Ricardo Daziano
daziano@cornell.edu

[1] Systems Science and Engineering, Cornell University, Ithaca, NY 14853, USA

[2] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA

Published online: 09 February 2022

paths. Another distinction that has been identified in the literature has to do with cycling experience. People that have less cycling experience also tend to have a stronger preference for segregation from motorized vehicles (Rossetti et al. 2018; Stinson and Bhat 2005). This information can be used by city planners to tailor their policies to the needs of different segments of the population.

The relationship between health and cycling has also been heavily studied, but unfortunately not from the point of view of infrastructure provision or preferences. The research questions relating the two have primarily focused on the effects cycling has on people's health. As expected, previous research has concluded that, on average, cyclists have a lower prevalence of diabetes, hypercholesterolemia, and obesity (Riiser et al. 2018; Lindström 2008). Understanding the interconnection of cycling preferences and health could lead to infrastructure that is better suited to the less healthy segment of the population, motivating this group to increase their cycling frequency and improve their health outcomes.

In this study, we address the relationship between self-assessed health status and infrastructure preferences. We do this using data collected from an online survey of New York City residents. We then use this data to estimate a latent class and latent variable choice model that describe health outcomes and cycling experience. Results show that respondents with higher body mass indices (BMI) and worse self-assessed health status have a stronger preference for segregated infrastructure and a lower sensitivity toward travel time.

The rest of the paper is organized as follows: The data collection process is presented first, with a description of the sample. Then, the latent class and latent variable methodology is described. After this, results are shown and discussed.

Data collection and preliminary analyses

We use microdata from an online survey carried out during December of 2019. This survey included several types of questions, including sociodemographic information, general travel patterns, and physical fitness indicators. Respondents were recruited from a representative Qualtrics panel. All respondents were regular New York City commuters (to work or school), over 18 years of age. Table 1 summarizes select characteristics of the sample.

The section of the survey that is most relevant to this study is a set of choice experiments regarding route choice using public bicycles. Each respondent faced seven binary choice scenarios, where two hypothetical routes were shown. The scenarios were described using words and images developed in a virtual city environment similar to a typical Manhattan avenue. Examples of the virtual cycling conditions are shown in Fig. 1, and the experimental attributes with their levels are shown in Table 2. An example of a choice card is shown in Fig. 2. A total of 5560 choices by 801 respondents were recorded.

Several effect indicators were also collected to identify respondents' health outcomes and cycling experience. We fitted a structural equation model to confirm the relationship between the effect indicators and the latent variables of interest, as well as to identify respondents' characteristics that correlate with the underlying factors. The significant indicators are shown in Table 3.

Table 1 Sample characteristics

Characteristic	Level	Value
Gender	Male	39.30%
	Female	60.70%
Age	Mean (std. dev.)	35.6 (13.7)
Household income	Less than \$10,000	5.90%
	\$10,000–\$15,000	2.60%
	\$15,000–\$25,000	7.70%
	\$25,000–\$35,000	8.90%
	\$35,000–\$50,000	10.90%
	\$50,000–\$75,000	19.70%
	\$75,000–\$100,000	15.60%
	\$100,000–\$150,000	11.20%
	\$150,000–\$200,000	5.10%
	\$200,000–\$500,000	4.70%
	More than \$500,000	2.00%
Race or ethnicity	American Indian or Alaska Native	< 0.1%
	Asian	12.60%
	Black or African American	25.60%
	Native Hawaiian or other Pacific Islander	< 0.1%
	White	46.20%
	Other, including multi-racial	0.10%
	Hispanic or Latino	28.00%
Cars available	None	34.70%
	One	47.40%
	Two	14.60%
	Three or more	3.20%
Home location	Bronx	15.20%
	Brooklyn	25.00%
	Manhattan	35.60%
	Queens	22.30%
	Staten Island	1.90%
BMI	Mean (std. dev.)	25.3 (5.9)
	Obese (BMI > 30)	18.70%
	Overweight (25 ≤ BMI < 30)	26.80%
	Healthy (18.5 ≤ BMI < 25)	53.80%
	Underweight (BMI < 18.5)	0.70%
Self-reported health	Excellent	25.47%
	Good	53.56%
	Average	19.73%
	Poor	1.25%
	Very poor	0.00%
Cycles at least once a week	During the fall	35.08%
	During the spring	38.58%

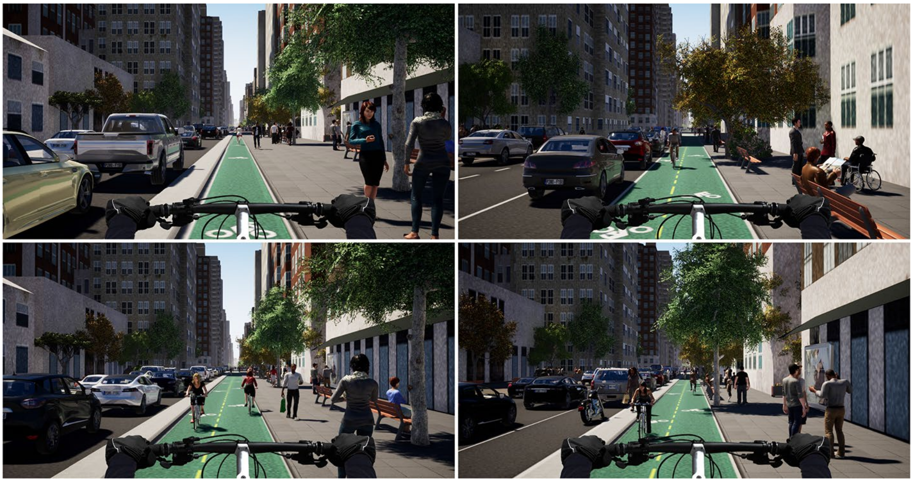

Fig. 1 Examples of choice scenarios presented to respondents

Table 2 Attribute levels of choice scenarios

Variable	Type	Levels
Travel time	Continuous	Pivoted around respondents' stated travel time.
Traffic/Speed	Categorical	Heavy traffic and slow speeds, or normal traffic flow with high speeds. This relationship was designed to replicate a slow, congested street, or an uncongested street with cars driving at the speed limit.
One- or two-way lane	Categorical	Either one- or two-way cycle lanes.
Parking	Categorical	Nonexistent, on left or on right.
Lane design	Categorical	Painted surface and/or with a buffer between the lane and cars. All choice scenarios had at least one of these possible protections.

The fitted structural equation model produced two underlying dimensions (latent variables): "experienced cyclist" and "poor health status." These, in turn, are negatively correlated between them (Fig. 3).

Methodology

To identify how preference structures vary across respondents depending on their general health outcomes, we use an integrated choice and latent class model. Nevertheless, because these health outcomes are not directly measurable using an online survey, we model them using latent variables. This produces an integrated choice, latent class and latent variable model. Each one of these components, as well as their integration, is described in the following subsections.

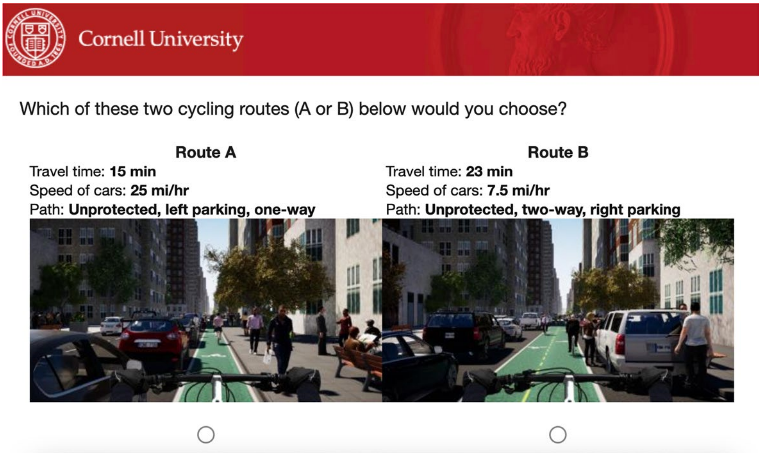

Fig. 2 Example of a choice card presented to a respondent

Table 3 Indicators used to fit a latent variable (LV) model using structural equation modeling

LV	Indicator	Type of response
Health outcomes		
	Body Mass Index (BMI)	Continuous. Constructed using stated height and weight.
	Self-reported health status	5 point Likert scale, from "Excellent" to "Very poor."
Cycling experience		
	Self-description of type of cyclist	4 point ordinal response, from "An advanced, confident cyclist who is comfortable riding in most traffic situations" to "I do not know how to bike."
	Uses app to access Citi Bike	Binary
	Bikes at least once a week during the fall or spring (two indicators)	Binary
	Typically walks or bikes during a weekday or weekend for more than 10 minutes (two indicators)	Binary

Latent class choice models

One strategy for modeling unobserved heterogeneity in preferences is to assume a discrete distribution of preferences, representing a discrete number of consumer categories or classes. Econometrically, the underlying categories may be inferred by estimating latent classes, as proposed by Kamakura and Russell (1989). Latent class choice models include two components: one relates individuals to the latent (unobserved) classes, whereas the other relates individuals to choices given their latent class.

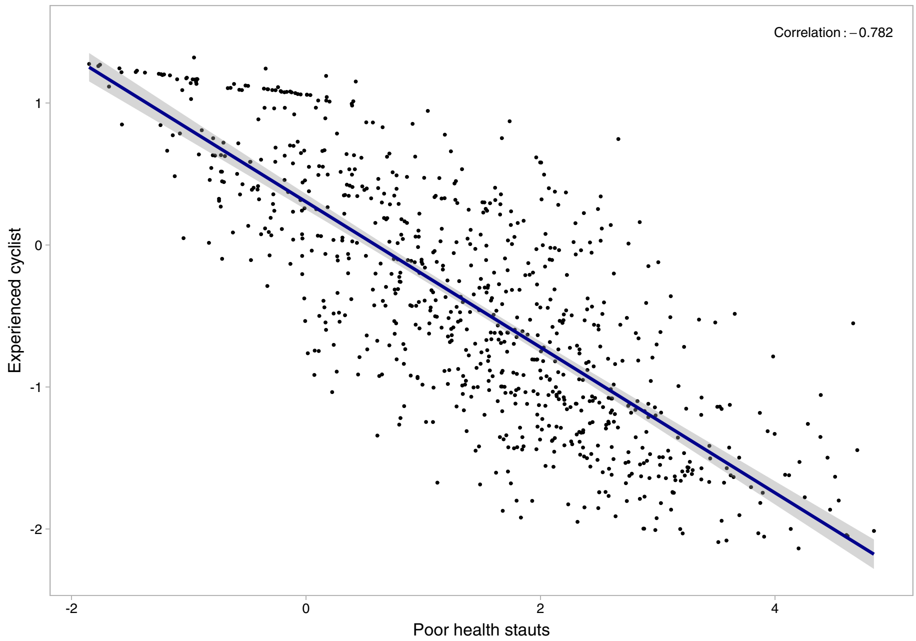

Fig. 3 Relation between the two latent variables produced by the structural equation model, at the respondent level

The utility derived by individual j when they choose alternative i given that they belong to class s can be represented by (1). $\mathbf{x}_{ij}$ is a vector of observed alternative attributes and consumer characteristics, and $\boldsymbol{\beta}^s$ is a vector of class-specific taste parameters. Utility U_{ij}^s can take different forms across classes, including varying distributional assumptions for the class-specific error component, ε_{ij}^s, and the specification of the indirect utility function, V^s.

$$U_{ij}^s = V^s\left(\mathbf{x}_{ij};\boldsymbol{\beta}^s\right) + \varepsilon_{ij}^s \tag{1}$$

If we assume, first, a random utility maximization framework and, second, that $\boldsymbol{\varepsilon}^s$ are independent and identically distributed Extreme Value Type I, then the probability that j chooses i given that they belong to class s is equal to the conditional logit choice probability (2). C_{js} is the choice set individual j faces given that they belong to class s in this equation. If V^s is assumed to have a linear specification, as is usually done in the literature, the scale parameter μ_s has to be normalized to ensure parameter identification.

$$P_j\left(i|s,\mathbf{x}_{ij};\boldsymbol{\beta}^s\right) = \frac{\exp\left(\mu_s V^s\left(\mathbf{x}_{ij};\boldsymbol{\beta}^s\right)\right)}{\sum_{l\in C_{js}} \exp\left(\mu_s V^s\left(\mathbf{x}_{lj};\boldsymbol{\beta}^s\right)\right)} \tag{2}$$

Since class membership cannot be directly observed, it is useful to construct some probabilistic measure relating individuals to classes. Let's define a class-membership link function W_{js} as shown in (3), where $\boldsymbol{\gamma}^s$ is a vector of class-specific parameters relating observable consumer characteristics, $\mathbf{x}_j$, with class s. Note that function Z may be specified in such a way that it only depends on a constant that must be estimated. Nevertheless, this approach is not informative on the relationship between individual characteristics and preference patterns.

$$W_{js} = Z\left(\mathbf{x}_j;\boldsymbol{\gamma}^s\right) + \zeta_{js} \tag{3}$$

We will assume the probability that a consumer j belongs to class s is proportional to the class-membership function W_{js}, and that $\boldsymbol{\zeta}$ are independent and identically distributed Extreme Value Type I. With this, the probability that j belongs to s is given by the multinomial logit probability (4). If Z has a linear specification, the scale parameter ς has to be once again normalized.

$$P_j\left(s|\mathbf{x}_j;\boldsymbol{\gamma}^s\right) = \frac{\exp\left(\varsigma \cdot Z\left(\mathbf{x}_j;\boldsymbol{\gamma}^s\right)\right)}{\sum_{p=1}^{S}\exp\left(\varsigma \cdot Z\left(\mathbf{x}_j;\boldsymbol{\gamma}^p\right)\right)} \tag{4}$$

To obtain the unconditional probability of j choosing i, we must marginalize $P_j(i|s)$ over $P_j(s)$, as shown in (5).

$$P_j\left(i|\mathbf{x}_{ij};\boldsymbol{\beta},\boldsymbol{\gamma}\right) = \sum_{s=1}^{S} P_j\left(i|s,\mathbf{x}_{ij};\boldsymbol{\beta}^s\right) \cdot P_j\left(s|\mathbf{x}_j;\boldsymbol{\gamma}^s\right) \tag{5}$$

One advantage this approach has is that it is fairly simple and straightforward. Moreover, since classes are discrete categories, this marginalization does not require to simulate an integral. This model's main disadvantage is that it is non-convex, which may make maximum likelihood estimation difficult.

The latent class logit model has been applied in varied settings. Some examples include preference for residential location (Walker and Li 2007), medical procedures (Ho et al. 2020; Rozier et al. 2019), transportation modes (El Zarwi et al. 2017; Hurtubia et al. 2014; Shen 2009; Bhat 1997), vehicle ownership (Ferguson et al. 2018), and in the field of environmental economics (Araghi et al. 2016; Beharry-Borg and Scarpa 2010).

The standard integrated choice and latent variable model (ICLV)

Another way of accounting for unobservable factors in the decision-making process is through latent variables. Latent variables are those that affect the decision-making process but cannot be directly measured. Previous research has used latent variables to model many kinds of qualitative constructs, including environmental concerns (Hess et al. 2013), risk aversion (Tsirimpa et al. 2010), or perceived quality (Palma et al. 2016).

A discrete choice model that considers unobservable attributes can be described by (6), where $\mathbf{x}_{ij}^*$ is a vector of latent variables. Assuming once again that $\boldsymbol{\varepsilon}$ are independent and identically distributed Extreme Value Type I and that V has a linear specification, the choice probability can be expressed as the conditional logit probability (7).

$$U_{ij} = V\left(\mathbf{x}_{ij},\mathbf{x}_{ij}^*;\boldsymbol{\beta}\right) + \varepsilon_{ij} \tag{6}$$

$$P_j\left(i|\mathbf{x}_{ij},\mathbf{x}^*_{ij};\boldsymbol{\beta}\right)=\frac{\exp\left(V\left(\mathbf{x}_{ij},\mathbf{x}^*_{ij};\boldsymbol{\beta}\right)\right)}{\sum_{l\in C_j}V\left(\mathbf{x}_{lj},\mathbf{x}^*_{lj};\boldsymbol{\beta}\right)} \tag{7}$$

To derive a choice probability that does not depend on unobservables, some distribution for the latent variable must be specified. This produces a stochastic relation between latent variables and observable variables, as shown in (8). Here, function X^* describes the structural relation between observable and unobservable variables through parameters $\boldsymbol{\lambda}$. The error term ω_{ij} accounts for variables not included in this model that affect $\mathbf{x}^*_{ij}$.

$$\mathbf{x}^*_{ij}=X^*\left(\mathbf{x}_{ij};\boldsymbol{\lambda}\right)+\omega_{ij} \tag{8}$$

The latent variable model is completed with a measurement relationship that can be expressed in general terms by (9), where function I relates the response to some effect indicator I_{ij} with the underlying latent construct $\mathbf{x}^*_{ij}$. Common specifications for these measurement relations are linear regressions when I_{ij} is continuous, or ordered logit or probit models when I_{ij} is a categorical variable, such as a Likert scale.

$$I_{ij}=I\left(\mathbf{x}^*_{ij};\boldsymbol{\tau}\right)+\nu_j \tag{9}$$

From this system of equations, the joint probability of observing j's choice and indicator I_{ij} unconditional on latent variables can be derived using (10), where g and f are density functions of I_{ij} and $\mathbf{x}^*_{ij}$ respectively.

$$P_j\left(i,I_{ij}|\mathbf{x}_{ij};\boldsymbol{\beta},\boldsymbol{\tau},\lambda\right)=\int P_j\left(i|\mathbf{x}_{ij},\mathbf{x}^*_{ij};\boldsymbol{\beta}\right)\cdot g\left(I_{ij}|\mathbf{x}^*_{ij};\boldsymbol{\tau}\right)\cdot f\left(\mathbf{x}^*_{ij};\lambda\right)d\mathbf{x}^*_{ij} \tag{10}$$

This now standard Integrated Choice and Latent Variable model (ICLV) was proposed by Walker and Ben-Akiva (2002) and has gained wide popularity in the choice modeling community, despite some criticisms (e.g., Chorus and Kroesen 2014; Kroesen and Chorus 2018). Even though most applications involve attitudinal latent variables (those that are related to some unobservable characteristic of consumers), perceptual latent variables can also be constructed (Bahamonde-Birke et al. 2015).

A latent class logit model with latent variables

An empirical problem of latent class choice models is that there is no clear interpretation of the fitted latent segments. What researchers usually do is to make intuitive sense of the overall segment by looking at the observable variables correlated with class membership model. These interpretations are hypotheses and not conclusions founded on the econometric model itself. If attitudinal latent variables are used to construct the class-membership model, direct and empirically well-founded relationships between latent constructs and class-specific taste parameters can be derived. This approach also frees the researcher from subjective interpretations of the parameters.

This method requires the integrated estimation of three components: a latent variable submodel, a latent class submodel, and a choice submodel. Assuming linear specifications and following the manner described above, each component can be described by the system of Eq. (11).

$$\text{Latent variable submodel} \qquad \mathbf{x}_j^* = \boldsymbol{\Lambda} \cdot \mathbf{x}_j + \boldsymbol{\omega}_j$$

$$I_j = I\left(\mathbf{x}_j^*; \boldsymbol{\tau}\right) + \nu_j$$

$$\text{Latent class submodel} \qquad W_{js} = \mathbf{x}_j^{*\prime} \boldsymbol{\gamma}^s + \zeta_{js} \quad \forall s \in \{1, \dots, S\}$$

$$P_j\left(s|\mathbf{x}_j^*; \boldsymbol{\gamma}\right) = \frac{\exp\left(\mathbf{x}_j^{*\prime} \boldsymbol{\gamma}^s\right)}{\sum_{p=1}^{S} \exp\left(\mathbf{x}_j^{*\prime} \boldsymbol{\gamma}^p\right)}$$

$$\text{Choice submodel} \qquad U_{ij}^s = \mathbf{x}_{ij}^{\prime} \boldsymbol{\beta}^s + \varepsilon_{ij}^s \quad \forall s \in \{1, \dots, S\}$$

$$P_j\left(i|s, \mathbf{x}_{ij}; \boldsymbol{\beta}^s\right) = \frac{\exp\left(\mathbf{x}_{ij}^{\prime} \boldsymbol{\beta}^s\right)}{\sum_{l \in C_{js}} \exp\left(\mathbf{x}_{lj}^{\prime} \boldsymbol{\beta}^s\right)} \quad \forall s \in \{1, \dots, S\}$$

$$P_j\left(i|\mathbf{x}_{ij}, \mathbf{x}_j^*; \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \sum_{s=1}^{S} P_j\left(i|s, \mathbf{x}_{ij}; \boldsymbol{\beta}^s\right) \cdot P_j\left(s|\mathbf{x}_j^*; \boldsymbol{\gamma}\right) \tag{11}$$

Each submodel is almost identical to the ones described above. In the case of the latent variable model, we now only identify user-specific latent variables since we are only concerned with attitudes. Since we allow more than one latent variable, the structural equation now includes matrix $\boldsymbol{\Lambda}$, which contains all $\boldsymbol{\lambda}$ vectors. The latent class submodel is now only specified as a function of the attitudinal latent variables $\mathbf{x}_j^*$, although observable user characteristics could be included as well. Finally, the choice component does not go through significant changes.

The joint probability of observing choice i and indicator I_j for individual j conditional on $\mathbf{x}_j^*$ is described by (12). As before, the unconditional probability (13) can be obtained by integrating over the distribution of $\mathbf{x}_j^*$.

$$P_j\left(i, I_j|\mathbf{x}_{ij}, \mathbf{x}_j^*; \boldsymbol{\beta}, \boldsymbol{\tau}, \lambda, \boldsymbol{\gamma}\right) = \left(\sum_{s=1}^{S} P_j\left(i|s, \mathbf{x}_{ij}; \boldsymbol{\beta}^s\right) \cdot P_j\left(s|\mathbf{x}_j^*; \boldsymbol{\gamma}\right)\right) \cdot g\left(I_j|\mathbf{x}_j^*; \boldsymbol{\tau}\right) \tag{12}$$

$$P_j\left(i, I_j|\mathbf{x}_{ij}; \boldsymbol{\beta}, \boldsymbol{\tau}, \lambda, \boldsymbol{\gamma}\right) = \int \left(\sum_{s=1}^{S} P_j\left(i|s, \mathbf{x}_{ij}; \boldsymbol{\beta}^s\right) \cdot P_j\left(s|\mathbf{x}_j^*; \boldsymbol{\gamma}\right)\right) \cdot g\left(I_j|\mathbf{x}_j^*; \boldsymbol{\tau}\right) \cdot f\left(\mathbf{x}_j^*; \lambda\right) d\mathbf{x}_j^* \tag{13}$$

Model parameters can be obtained using maximum likelihood estimation. Assuming that there are a total of J respondents and that each respondent j observed T_j choice scenarios, the likelihood can be expressed as:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda, \boldsymbol{\gamma}|\mathbf{x}) = \prod_{j=1}^{J} \int \prod_{t=1}^{T_j} P_j\left(i|\mathbf{x}_{ijt}, \mathbf{x}_j^*; \boldsymbol{\beta}, \boldsymbol{\gamma}\right) \cdot g\left(I_{ij}|\mathbf{x}_j^*; \boldsymbol{\tau}\right) \cdot f\left(\mathbf{x}_j^*; \lambda\right) d\mathbf{x}_j^* \tag{14}$$

There are a few examples of this model being used in the literature, including Hess et al. (2013) and Krueger et al. (2018).

Results

The following subsections discuss the results of modeling the data presented in a previous section using the latent class and latent variable method. We will first discuss direct estimates, and then analyze marginal rates of substitution of the two models obtained. All results shown were obtained using the `apollo` package in R (Hess and Palma 2019).

Main results

This section presents results for two latent class and latent variable models. The one that addresses this study's research question uses a latent variable that describes how health status correlates with infrastructure preferences. The second one relates cycling experience to these preferences. The latter model was estimated to compare and validate the results of the former one. Note that because these two latent variables are highly correlated (see Fig. 3), both could not be integrated into a single model. Finally, a standard conditional logit was also estimated to have a baseline comparison for parameter estimates, marginal rates of substitution, and goodness-of-fit measures. Table 4 shows the results for all models.

The general structure of both latent class choice models with latent variables is shown in Fig. 4. In both cases, sociodemographic variables inform the structural component of the latent variable submodel. These latent variables then inform class membership, which in turn define the values of the utility functions' parameters in the choice component.

First, the likelihood values at convergence of the choice components for the latent class and latent variable models are higher than the one for the baseline MNL model. This result shows that there is significant preference heterogeneity that cannot be captured by the conditional logit. The likelihoods of the choice components of the two latent class and latent variable models are similar, showing that both provide a meaningful way to segment the population. During model development, we also verified that standard ICLV models provided worse fit and poorer behavioral information, which is why we decided to only report the models in Table 4.

The latent variable model shows that "Poor health status" and "Experienced cyclist" tend to have parameters with opposite signs, which is consistent with the negative correlation found using structural equation modeling. Results show that respondents tend to have better health status and cycling experience if they are men, younger, own a car, and live in Manhattan, as opposed to other New York City boroughs. Some of these results are consistent with previous findings. For example, Rossetti et al. (2018) also found that younger men tend to be more experienced cyclists.

People that have a better health status and more experience cycling have a higher probability of belonging to Class 1 of Model 1 and Class 2 of Model 2. These classes show similar preference structures. For example, both have a negative parameter related to travel time, as expected. Moreover, these individuals show distaste for parking, and preference for painted and buffered cycle lanes. These results are in line with previous findings for people that have cycling experience (e.g., Rossetti et al. 2018; Stinson and Bhat 2005). Class 1 of Model 1 and Class 2 of Model 2 also have the same signs as the parameters in the baseline MNL model.

There are fewer significant taste parameters in Class 2 of Model 1 ("Worse health status") and Class 1 of Model 2 ("Less experience cycling"). In both cases, there is no significant effect of traffic volume and speed, parking, and only a slightly significant and

Table 4 Integrated choice, latent class and latent variable models, together with a baseline multinomial logit model (MNL)

		Baseline: MNL	Model 1: Poor health stat.	Model 2: Exp. cyclist
Choice component				
Class 1			*Better health*	*Less experienced*
Time		−0.0389*** (−5.41)	−0.0636*** (−6.80)	1.34*(2.05)
Traffic	Normal	0 (fixed)	0 (fixed)	0 (fixed)
	Heavy	0.174*** (3.67)	0.116 (1.79)	0.211 (0.76)
One- or two-way	One-way	0 (fixed)	0 (fixed)	0 (fixed)
	Two-way	0.299*** (6.79)	0.163** (2.86)	5.02* (2.08)
Parking	Nonexistent	0 (fixed)	0 (fixed)	0 (fixed)
	On left	−0.770*** (−5.42)	−0.620*** (−3.34)	−3.45 (−1.00)
	On right	−0.261*** (−5.04)	−0.415*** (−6.10)	−5.94 (−1.47)
Lane design†	Paint	0.294* (2.21)	0.582*** (3.77)	10.0*** (11.24)
	Buffer	1.33*** (22.45)	0.695*** (10.35)	13.5* (2.06)
Class 2			*Worse health*	*More experienced*
Time			1.11* (2.18)	−0.0635*** (−6.94)
Traffic	Normal		0 (fixed)	0 (fixed)
	Heavy		0.138 (0.54)	0.111 (1.74)
One- or two-way	One-way		0 (fixed)	0 (fixed)
	Two way		4.19* (2.25)	0.164** (2.91)
Parking	Nonexistent		0 (fixed)	0 (fixed)
	On left		−1.74 (−0.69)	−0.622*** (−3.37)
	On right		−4.18 (−1.46)	−0.418*** (−6.26)
Lane design†	Paint		12.1*** (17.38)	0.583*** (3.82)
	Buffer		10.8* (2.28)	0.699*** (10.41)
Class membership component (Class 2)				
Intercept			−0.653*** (−3.38)	0.546*** (4.11)
Poor health status			0.487* (2.51)	
Exp. cyclist				0.827*** (3.33)
Latent variable component				
Female			0.625*** (4.11)	−0.202*** (−4.69)
Age			0.0192*** (4.00)	−0.00345* (−2.29)
Driver's license			−0.497* (−2.55)	0.110* (2.22)
Cars	None		0 (fixed)	0 (fixed)
	One		−0.271 (−1.54)	0.106* (2.17)
	Two or more		−0.696** (−2.89)	0.142* (2.44)
Employed			−0.496** (−2.78)	0.0587 (1.38)
Home location	Manhattan		0 (fixed)	0 (fixed)
	Bronx		0.509 (1.88)	−0.0762 (−1.15)
	Brooklyn		0.413* (2.21)	−0.135** (−2.68)
	Queens		0.530** (2.63)	−0.175** (−3.22)
Standard deviation			1.11*** (3.62)	0.436*** (12.15)
# Of individuals		801	801	801
# Of observations		5560	5,60	5560

Table 4 (continued)

	Baseline: MNL	Model 1: Poor health stat.	Model 2: Exp. cyclist
Log-likelihood	−2898.89	−6056.47	−7094.20
Log-likelihood (choice)	−2898.89	−2701.53	−2709.00
Draws	–	1000 (Halton)	1000 (Halton)
# Of parameters	7	30	30

$*** \ p < 0.001$, $** \ p < 0.01$, $* \ p < 0.05$. †All choice scenarios had at least one of these levels

Robust std. errors reported. Parameters of measurement eqs. not reported

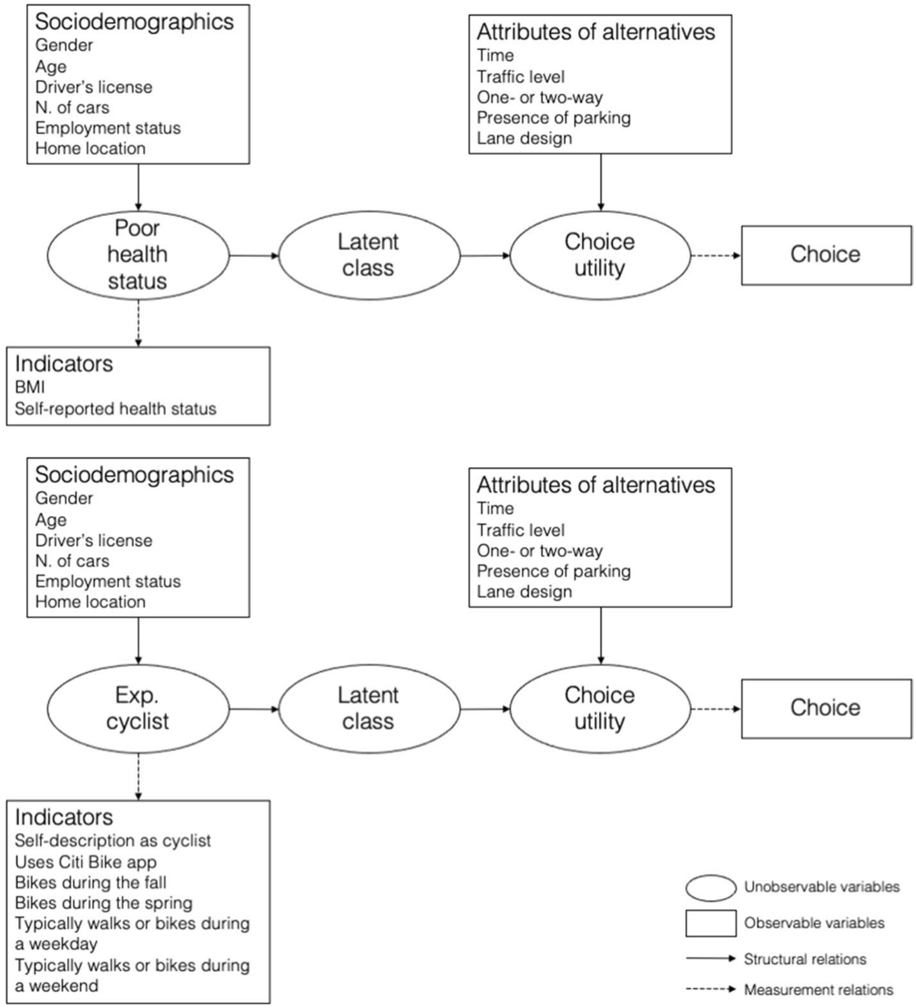

Fig. 4 Structure of latent class choice models with latent variables. Top: Model that includes "Poor health status." Bottom: Model that includes "Experienced cyclist"

positive effect of two-way lanes. Painted and buffered lanes have a strong positive effect on choice.

Contrary to economic theory, these classes have positive time parameters, which means that members of these classes prefer longer routes when all else is equal. This result is counterintuitive and may reflect a proxy for noncompensatory valuation of travel time. Another source of error could be that individuals from these classes don't experience cycling as often, and therefore may not have adequately assessed the choice experiments. Unfortunately, given the adopted design and data, it was not possible to disentangle such issue. Finally, nobody in the sample had a class-membership probability close to one or zero (see Fig. 5), which means that we do not have evidence of an individual with a net-positive travel time parameter. Therefore, conclusions derived from this result should be analyzed with care.

Marginal rates of substitution

The ideal bicycle lane design has been a matter of debate among urban designers. City planners usually have to deal with the trade-off between segregation from cars–something the literature has consistently demonstrated is desirable for cyclists–and cost. Whereas cheaper bicycle lanes allow to expand the network at a faster pace, this cheaper infrastructure can fail to attract new riders. Given this dichotomy, the marginal rate of substitution (MRS) between different kinds of designs and travel time can help assess the costs and social benefits of different approaches to cycling infrastructure provision.

Since the results for the classes with worse health outcomes and less cycling experience had counterintuitive results, we will concentrate our analysis on the classes that did follow a pattern consistent with the literature and economic theory.

Table 5 shows MRSs of interest. Since the maximum likelihood estimation parameters asymptotically distribute Normal, the MRSs were derived using the Delta method (Daly et al. 2012).

As expected, the class of experienced cyclists and people with better health outcomes are willing to increase their travel time in exchange of a more protected cycle lane. The detours these individuals were willing to make also fall within reasonable ranges: on average, members of these classes are willing to increase their travel time by approximately 9 and 11 minutes to access a painted and buffered cycle lane, respectively. The MRSs

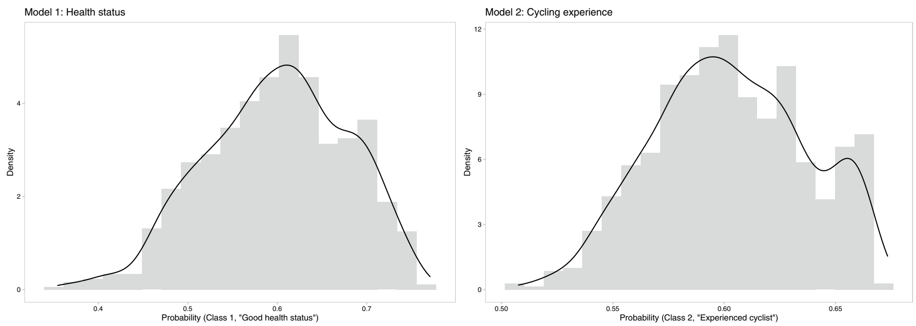

Fig. 5 Empirical distributions of class-membership probabilities

Table 5 Select marginal rates of substitution (MRS)

	Baseline: MNL	Model 1: Health status Class 1	Model 2: Cycling experience Class 2
Time and Paint	-7.56^{*} min. (−2.05)	-9.15^{**} min. (−3.23)	-9.17^{***} min. (−3.30)
Time and Buffer	-34.23^{***} min. (−5.75)	-10.94^{***} min. (−6.81)	-11.00^{***} min. (−7.09)
Paint and Buffer†	0.221^{***} (−7.51)	0.836 (−0.66)	0.834^{***} (3.39)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.5$. MRS and (t-satistics) reported.

Robust standard errors used for Delta method. †t-statistics calculated with respect to 1

between the two types of design show that the differences between both are either not significant, or that buffered lanes are preferable.

While the MRS between time and "Paint" obtained using the multinomial logit model (MNL) was similar to the ones obtained using the latent class models, the other two were quite different. For example, while the latent classes we analyzed were willing to increase their travel time on average by 11 minutes to access a buffered lane, the conditional logit model predicts this number is around 34 minutes. The differences between both models could be due to the MNL's inability to capture underlying heterogeneity.

Conclusions

Previous research dedicated to identifying preferences for cycling infrastructure has failed to consider the relationships of those preferences with health status. Understanding this association is essential for policymakers to improve health outcomes from the low-impact physical exercise that comes from cycling. If the specific needs of those with poorer health outcomes are addressed in the infrastructure design process, there is a higher likelihood that they will engage in active transportation and improve their health.

We used a stated preference data set from New York City to fit an integrated choice, latent class and latent variable model to identify the relations between health and infrastructure preference. Results show, first, that experienced cyclists have similar taste patterns as people in good health. Second, we found two archetypal taste structures, defined by latent classes, between which respondents lie. Third, we found that the unhealthier latent class had a positive time parameter, which could indicate non-compensatory behavior or an inability to adequately respond to the choice scenarios due to cycling inexperience.

This study provides evidence that supports a double benefit from policies that promote cycling among the inexperienced. Since the inexperienced and the less healthy have similar taste structures, cycling infrastructure designs that appeal to the former will likely appeal to the latter, and vice versa. Whenever this is the case, there is a potential benefit of producing a shift towards more sustainable modes of transportation, thus alleviating congestion and carbon emissions, and promoting more physical exercise among the population that is less physically fit. This double benefit has the potential to reduce public health spending, as well as to decrease future spending to counter the effects of climate change.

Acknowledgements This research was supported by the Center for Transportation, Environment, and Community Health, CTECH (data collection); the National Science Foundation Award No. SES-2031841 (methodology); and the Chilean National Agency for Research and Development's Scholarship Program (ANID) /

Doctorado Becas Chile / 2019 / 72200167 (graduate research assistant support). On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author contributions Project conceptualization was done by both authors. Material preparation and data collection was performed by Ricardo A. Daziano. Model estimation was performed by Tomás Rossetti. A first draft of the manuscript was written by Tomás Rossetti and Ricardo A. Daziano edited it. All authors read and approved the final manuscript.

References

Aldred, R., Elliott, B., Woodcock, J., Goodman, A.: Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age. Transp. Rev. **1647**, 1–27 (2016)

Araghi, Y., Kroesen, M., Molin, E., Van Wee, B.: Revealing heterogeneity in air travelers' responses to passenger-oriented environmental policies: a discrete-choice latent class model. Int. J. Sustain. Transp. **10**(9), 765–772 (2016)

Bahamonde-Birke, F., Kunert, U., Link, H., Ortúzar, J.D.D.: About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. Transportation **42**(6), 1–19 (2015)

Beharry-Borg, N., Scarpa, R.: Valuing quality changes in Caribbean coastal waters for heterogeneous beach visitors. Ecol. Econ. **69**(5), 1124–1139 (2010)

Bhat, C.R.: An endogenous segmentation mode choice model with an application to intercity travel. Transp. Sci. **31**(1), 34–48 (1997)

Buehler, R., Dill, J.: Bikeway networks: a review of effects on cycling. Transp. Rev. **36**(1), 9–27 (2016)

Chorus, C.G., Kroesen, M.: On the (im-)possibility of deriving transport policy implications from hybrid choice models. Transp. Policy **36**, 217–222 (2014)

Daly, A., Hess, S., de Jong, G.: Calculating errors for measures derived from choice modelling estimates. Transp. Res. Part B Methodol. **46**(2), 333–341 (2012)

El Zarwi, F., Vij, A., Walker, J.L.: A discrete choice framework for modeling and forecasting the adoption and diffusion of new transportation services. Transp. Res. Part C Emerg. Technol. **79**, 207–223 (2017)

Ferguson, M., Mohamed, M., Higgins, C.D., Abotalebi, E., Kanaroglou, P.: How open are Canadian households to electric vehicles? A national latent class choice analysis with willingness-to-pay and metropolitan characterization. Transp. Res. Part D Trans. Environ. **58**(2017), 208–224 (2018)

Hess, S., Palma, D.: Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. J. Choice Model. **32**, 1–43 (2019)

Hess, S., Shires, J., Jopson, A.: Accommodating underlying pro-environmental attitudes in a rail travel context: application of a latent variable latent class specification. Transp. Res. Part D Transp. Environ. **25**, 42–48 (2013)

Ho, K.A., Acar, M., Puig, A., Hutas, G., Fifer, S.: What do Australian patients with inflammatory arthritis value in treatment? A discrete choice experiment. Clin. Rheumatol. **39**(4), 1077–1089 (2020)

Hurtubia, R., Nguyen, M.H., Glerum, A., Bierlaire, M.: Integrating psychometric indicators in latent class choice models. Transp. Res. Part A Pol. Prac. **64**, 135–146 (2014)

Kamakura, W., Russell, G.: A probabilistic choice model for market segmentation and elasticity structure. J. Mark. Res. **26**(4), 379–390 (1989)

Kroesen, M., Chorus, C.: The role of general and specific attitudes in predicting travel behavior-a fatal dilemma? Travel Behav. Soc. **10**, 33–41 (2018)

Krueger, R., Vij, A., Rashidi, T.H.: Normative beliefs and modality styles: a latent class and latent variable model of travel behaviour. Transportation **45**(3), 789–825 (2018)

Lindström, M.: Means of transportation to work and overweight and obesity: a population-based study in southern Sweden. Prev. Med. **46**(1), 22–28 (2008)

Nello-Deakin, S.: Environmental determinants of cycling: Not seeing the forest for the trees? J. Transp. Geogr. **85**(2019), 102704 (2020)

Palma, D., Ortúzar, J., d. D., Rizzi, L. I., Guevara, C. A., Casaubon, G., and Ma, H.: Modelling choice when price is a cue for quality a case study with Chinese wine consumers. J. Choice Model. **19**, 24–39 (2016)

Pucher, J., Buehler, R.: Making cycling irresistible: lessons from the Netherlands. Denmark and Germany. Transp. Rev. **28**(4), 495–528 (2008)

Riiser, A., Solbraa, A., Jenum, A.K., Birkeland, K.I., Andersen, L.B.: Cycling and walking for transport and their associations with diabetes and risk factors for cardiovascular disease. J. Transp. Health **11**(2017), 193–201 (2018)
Rossetti, T., Guevara, C.A., Galilea, P., Hurtubia, R.: Modeling safety as a perceptual latent variable to assess cycling infrastructure. Transp. Res. Part A Policy Prac. **111**(February), 252–265 (2018)
Rozier, M.D., Ghaferi, A.A., Rose, A., Simon, N.J., Birkmeyer, N., Prosser, L.A.: Patient preferences for bariatric surgery: findings from a survey using discrete choice experiment methodology. JAMA Surg. **154**(1), 1–10 (2019)
Shen, J.: Latent class model or mixed logit model? A comparison by transport mode choice data. Appl. Econ. **41**(22), 2915–2924 (2009)
Stinson, M. A., Bhat, C. R.: A comparison of the route preferences of experienced and inexperienced bicycle commuters. Transportation Research Board 84th Annual Meeting, (512), (2005)
Tsirimpa, A., Polydoropoulou, A., Antoniou, C.: Development of a latent variable model to capture the impact of risk aversion on travelers' switching behavior. J. Choice Model. **3**(1), 127–148 (2010)
Walker, J., Ben-Akiva, M.: Generalized random utility model. Math. Soc. Sci. **43**(3), 303–343 (2002)
Walker, J., Li, J.: Latent lifestyle preferences and household location decisions. J. Geogr. Syst. **9**, 77–101 (2007)