

A Comprehensive Data Fusion to Evaluate the Impacts of COVID-19 on Passenger Travel Demands

Application of a Core-Satellite Data Collection Paradigm

Sanjana Hossain, Patrick Loa, Kaili Wang, Sk. Md. Mashrur, Alireza Dianat, Khandker Nurul Habib

January 11, 2023



UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

Overview

1. Background
2. Study design
3. Data fusion method
4. Results
5. Application of fused data for choice modelling
6. Conclusion

Background



Attitudes and perceptions are important predictors of travel choices



Traditional household travel surveys rarely include attitudinal questions



Specialized surveys can collect detailed attitudinal data but suffer from limited sample sizes



Need a method to fuse “core” household travel survey with “satellite” surveys that collect attitudinal data

Research Objective



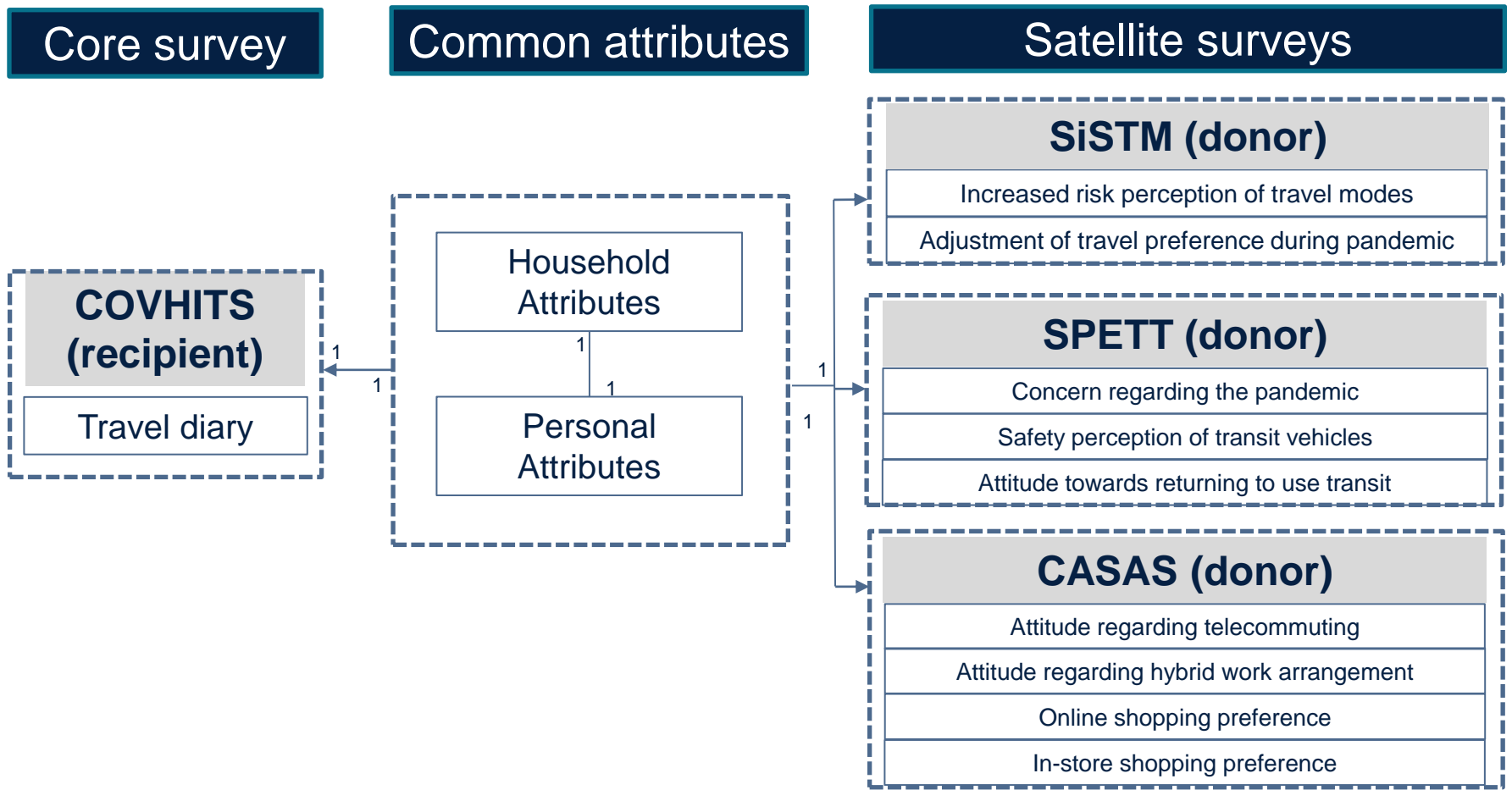
Propose a data fusion method to enrich a “core” household travel survey by linking it to three “satellite” surveys that collected rich attitudinal information



Study area: Greater Toronto Area

	Survey	Description	Study period	Sample size
Core	COVHITS	CO Vid-19 influenced H ouseholds' Interrupted T ravel S chedule	Oct - Nov' 21	8,911 individuals
Satellites	SiSTM	S tudy into the use of S hared T ravel M odes	Jul' 21	767 individuals
	SPETT	S tated P reference E xperiment on T ravel mode and especially T ransit choice behavior	Jul' 21	849 individuals
	CASAS	C ovid A ctivity S cheduling and A daptation S urvey	Jul' 21	860 individuals

Core & satellite survey design



Data fusion method

Step 1: Harmonisation and reconciliation of sources

- Make common variables coherent in attribute levels
- Ensure that the samples refer to the same population

Step 2: Analysis of the explanatory power for common variables

- Cramer's V used to analyze association of common variables with target variables
- Common variables with good explanatory power selected as matching variables

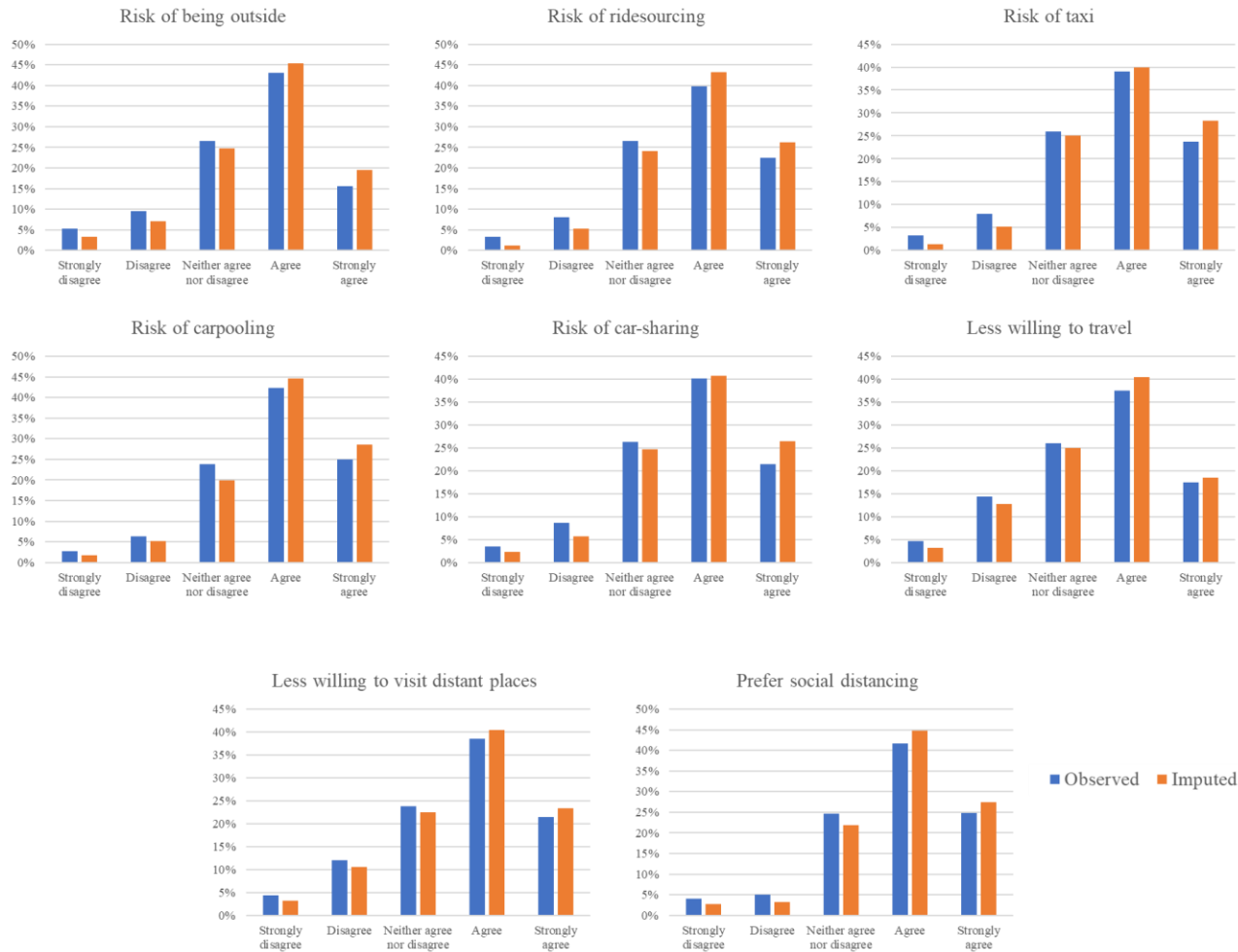
Step 3: Matching method

- k-NN type extension of the hot-deck imputation technique
- Generate multiple instances of fused dataset using Monte Carlo draws from the k NN
- Gower's dissimilarity coefficient used as the measure of distance

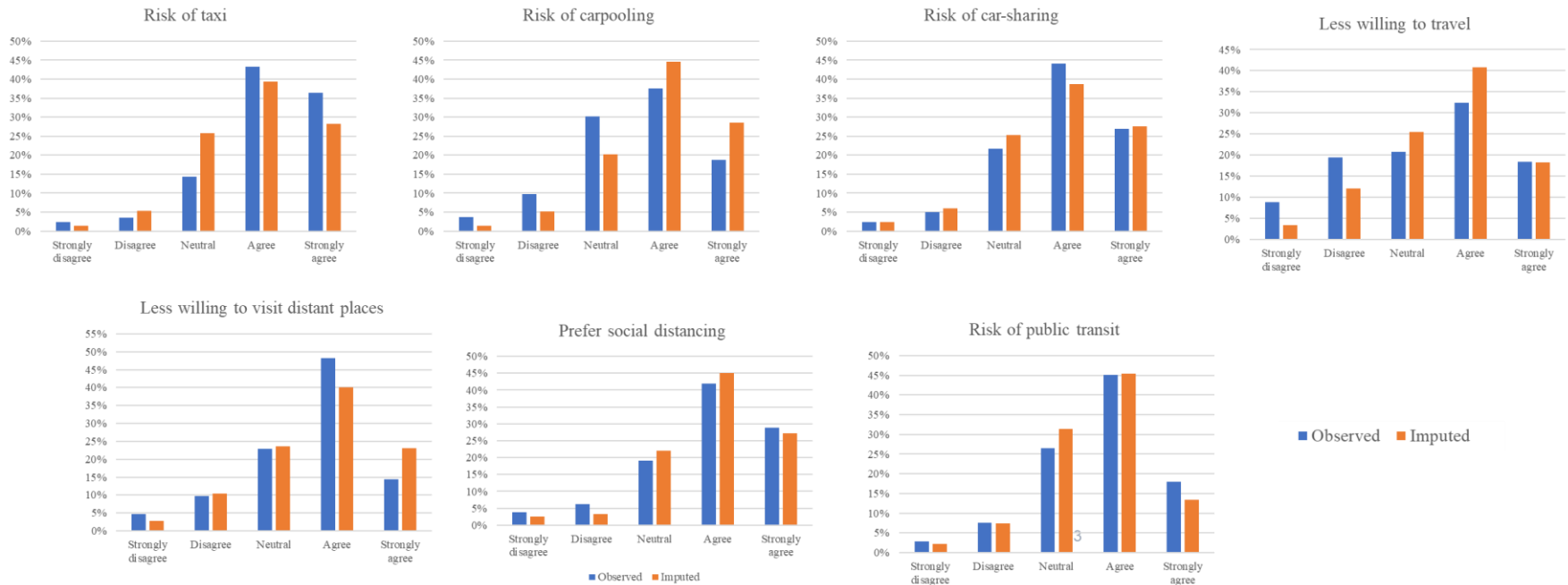
Step 4: Quality evaluation

- Compare marginal distributions of target variables in donor and the fused datasets
- Compare true (observed) values of target variables of the actual recipient units in the core survey with their imputed values

Preservation of marginal distributions in satellite (donor) and fused data



Validation of data fusion outputs



- Fusion replicates general trend of most attitudinal responses
- For some variables like “risk of carpooling” and “less willing to travel”, the distributions are somewhat different, indicating that people are gradually getting used to the pandemic

Data fusion results – socioeconomic attributes



Imputed attitudinal variables meet a priori expectations regarding socioeconomic status of respondents



Age is a significant factor affecting individuals' perception of risks and adjustment to travel during the pandemic



Older respondents have higher risk perception, are more concerned about the pandemic, perceive public transit as less safe, and have a greater preference for in-store shopping

Data fusion results – travel behaviour



Individuals with higher imputed levels of perceived risks made fewer trips



Individuals who agree with the advantages of telecommuting completed fewer work trips per day



Individuals who prefer online grocery shopping made fewer shopping trips, and vice versa

Data fusion results – travel behaviour



Individuals with higher perceived risk of pandemic rely more on driving and avoid public transit



Similarly, individuals who adjusted their travel patterns during this period rely more on driving



Among the different types of transit vehicles, bus/streetcar is perceived to be the least safe

Application of fused data for choice modelling

- Empirical investigation conducted with the synthetic fused data
- Demonstrate how to use the fusion outputs for subsequent modelling
- Hybrid commute mode choice model estimated with the fused data
 - a subset of the travel diary data representing commuting trips from the core survey
 - the socio-demographic information of the respondents
 - their attitudinal statements imputed from the satellite surveys

Hybrid choice model estimation

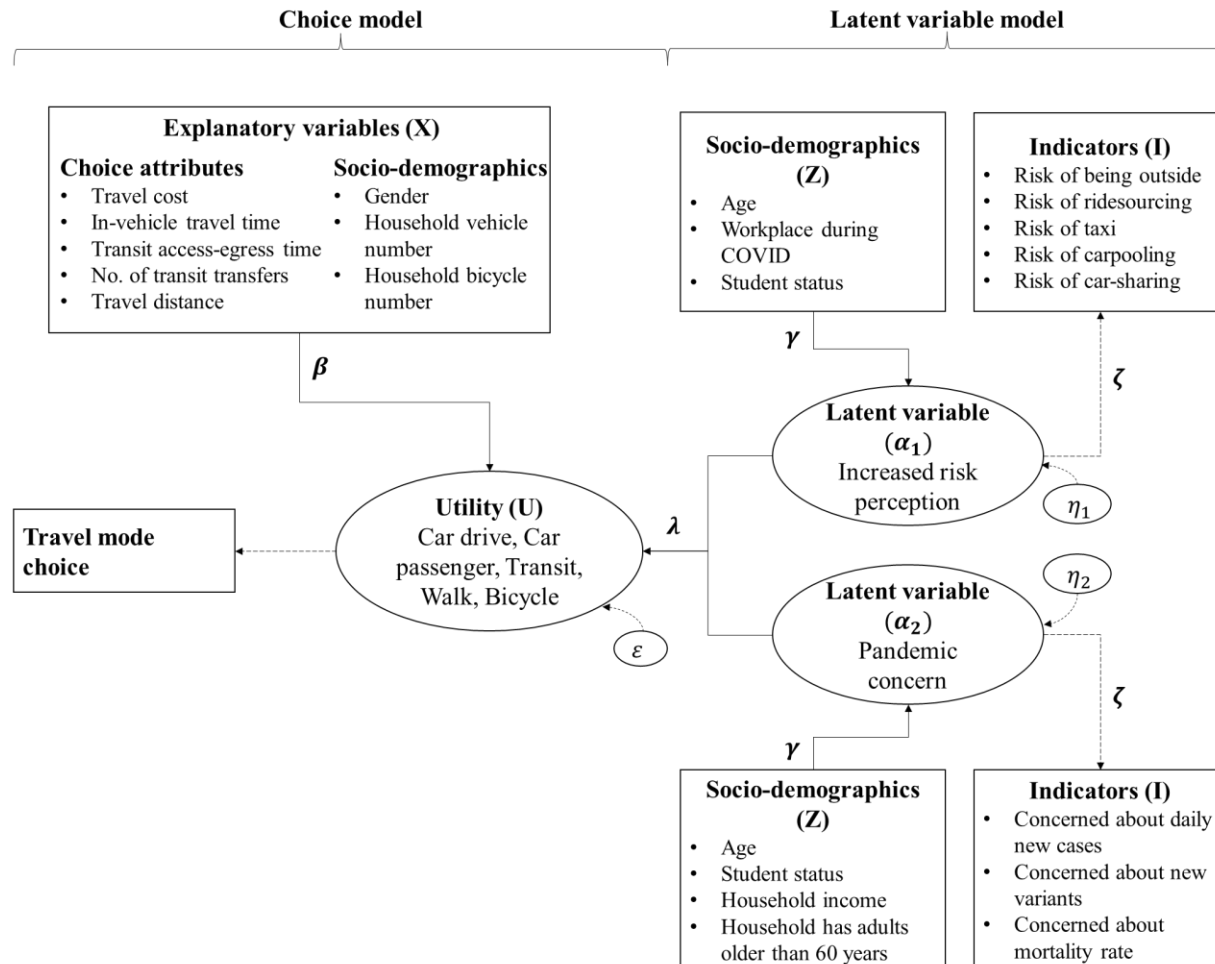
- Five major commute modes: car drive, car passenger, transit, walk, and bicycle
- Transportation level-of-service (LOS) attributes
 - Travel time generated using Google directions API
 - Auto cost generated using cost matrices widely used for transportation planning in the study region
 - Transit fare generated a calibrated Deterministic User Equilibrium traffic assignment model of the study area called the GTA model was used
- Model estimated using each of the synthetic fused datasets

Hybrid choice model specification: Factor analysis

- Factor analysis: to identify latent factors based on the imputed attitudinal questions
- Consistent findings obtained using two factors (with loadings larger than 0.4)

Latent Construct	Observed indicator	Factor Loading
Perception of increased risk during the pandemic	I believe there are more risks associated with leaving my home than before the pandemic	0.402
	I believe there is more risk associated with using ride-sourcing services than before the pandemic	0.400
	I believe there is more risk associated with using taxi services than before the pandemic	0.494
	I believe there is more risk associated with carpooling than before the pandemic	0.445
	I believe there is more risk associated with using car-sharing services (e.g., Zipcar, Communauto) than before the pandemic	0.436
Concerns regarding the pandemic	I am concerned about the number of daily new cases in Ontario, Canada	0.479
	I am concerned about the emergence of the new variant of COVID-19	0.483
	I am concerned about the mortality rate of the disease which is causing the pandemic	0.445

Final hybrid choice model specification



Choice model results



LOS attributes (travel cost, trip length, different travel time components, number of transit transfers) have –ve signs



Females are less likely to cycle than males



Household vehicle and bicycle ownership positively affect car use (car drive and car passenger) and bicycle use

Choice model results – latent attitudes



“Increased risk perception” has +ve effect on car drive mode and negative effect on shared ride mode



Individuals who have higher “pandemic concern” are less likely to choose transit for commuting

Structural and measurement models results



Older respondents and respondents who had to be physically present in their workplace during the pandemic have higher risk perceptions



Respondents who had to be physically present in their workplace during the pandemic have higher risk perceptions



Older respondents and respondents who lived with senior household members have increased pandemic concern



Individuals whose household income is below \$60,000 are less likely to be concerned about the pandemic than higher-income individuals

Key findings

- The study presents a proof of concept of how the implicit data fusion method may be used to integrate multiple travel survey data
- The fused data can be reliably used for much more complex and stable investigations than would be possible individually with either the core or the satellite survey data
- Imputing multiple fused datasets helps reduce potential biases that can affect subsequent analyses using the data
- Ideal satellite design should ensure
 - Comprehensive set of consistent, coherent common variables that are well associated with the target variables
 - Same survey conduction period (to control for any external effects)

Questions?

sanjana.hossain@mail.utoronto.ca

patrick.loa@mail.utoronto.ca