The Evolution Platform: Enhancing Travel Survey Data Quality Through Paradata Analysis and Respondent Behavior Monitoring

Geneviève Bastien Pierre-Léo Bourbonnais Yannick Brosseau

Samuel Duhaime-Morissette Catherine Morency

February 2025

1 Introduction

High-quality data collection is crucial for transportation planning and policy development. Household travel surveys serve as key instruments for collecting demographic, trip, mode choice, spatial, and temporal data essential for informed decision-making. However, collecting reliable transportation data presents numerous challenges, including respondent fatigue, recall errors, and complexity in capturing accurate spatial information.

Traditional travel surveys conducted using pen-and-paper trip diaries or Computer-Assisted Telephone Interviewing (CATI) have evolved in response to declining response rates and increasing smartphone availability. While there is abundant literature on the effects of specific survey methods on response rates, most studies rely on custom, small or medium-scale, one-time experiments that do not propose proper tools to implement future versions of those surveys (Verzosa et al., 2021; Bayart and Bonnel, 2015).

This paper presents the Evolution open-source platform, designed to facilitate high-quality travel surveys with a strong emphasis on data quality control and respondent behavior analysis. The platform integrates multiple data collection modes, systematic validation procedures, and comprehensive paradata collection (Vehovar et al. (2024)) to improve both the data collection process and final data quality.

2 Challenges in Survey Data Collection

Household travel surveys face several data quality challenges. Respondents must recall their trips accurately and may report trips for other household members, generating proxy respondent bias. Respondent fatigue leads to lower data quality, increased missing trips, or soft refusals (Armoogum et al., 2024). Web survey respondents tend to omit short walking trips, further compromising data quality (Bayart and Bonnel, 2024; Colaço and de Abreu e Silva, 2024).

Available survey tools often fail to address these challenges adequately. Commercial platforms like Qualtrics offer strong performance but require custom development for geospatial awareness that cannot be shared with the community. Open-source options like LimeSurvey, KoboToolbox, and SurveyJS lack support for complex survey flows, paradata collection, and comprehensive validation capabilities essential for quality travel surveys.

These limitations have led many organizations to develop in-house survey applications. Several such developments exist in the literature, including TRAISI (Chung et al., 2017), TSaaS (Vardhan et al., 2022), and an interactive trip diary (Barthelmes et al., 2024). However, many lack proper maintenance, documentation, or licensing, limiting their broader application and community development.

3 The Evolution Platform

Evolution was designed to maximize data quality while minimizing respondent burden. The platform provides integrated tools for monitoring the data collection process, analyzing respondent behavior through paradata, and enhancing survey design based on empirical evidence. (Evolution (2025))

3.1 Quality-Focused Features

Evolution implements several key features to ensure high-quality data collection:

Multi-modal data collection: The platform supports both self-administered (CAWI) and interviewer-aided (CATI/CAPI) surveys. This integration allows researchers to combine different data collection methods, ensuring comprehensive coverage and reducing mode-specific biases.

Geography-aware trip diary: Evolution's trip diary interface captures spatial data in two steps: 1) declaration and geocoding of all visited places with arrival and departure times, and 2) transportation modes used for each trip. This structured approach enhances the accuracy of spatial and temporal data.

Real-time validation: The platform implements comprehensive validation checks during data entry, allowing respondents to correct inconsistencies immediately. Integration with routing engines enables detection of trips with implausible speeds or impossible routes.

Customizable questionnaire flow: Dynamic question appearance, labeling, and response choice filtering based on previous answers ensures respondents only see relevant questions and options, reducing burden and enhancing data relevance.

Audit and review system: A robust post-collection auditing system flags potential errors, inconsistencies, or outliers for further review, enhancing final data quality.

3.2 Respondent Behavior Analysis Through Paradata

A significant innovation in Evolution is its comprehensive paradata collection framework. Paradata includes response times, answer modifications, page navigation patterns, and device information. This data provides insights into respondent behavior and questionnaire effectiveness:

Response time tracking: Timestamps for each question reveal which items require more cognitive effort or cause confusion, informing questionnaire improvements.

Answer modification monitoring: The platform records when respondents change answers, indicating question ambiguity or comprehension issues.

Navigation patterns: Page visit sequences and durations help identify questionnaire flow issues and drop-off points.

Device usage analysis: Data on device types and their impact on completion rates enables interface optimization for all platforms.

These paradata elements allow researchers to analyze the survey instrument itself, leading to continuous improvement in questionnaire design and implementation. For example, Keshavarz Mohammadian (2023) found that 84% of respondents who completed a survey using Evolution under-estimated its duration and perceived it as shorter than it actually was, suggesting effective user experience design.

4 Case Studies and Empirical Evidence

Evolution has been used in various surveys ranging from academic research to large-scale regional travel surveys. Table 1 summarizes key metrics from selected surveys conducted with the platform over the past five years.

Survey	Sample size	Completion rate	Avg. duration (min)
COVID impact (2020-2022)	7,059	60.6%	22.5
Montreal regional (2023)	146,673	76.2%	18.9
Quebec City regional (2023)	51,304	72.7%	16.9
Long-distance travel (2024)	11,066	62.4%	9.0

Table 1: Summary of surveys conducted using the Evolution platform

4.1 Device Impact on Data Quality

Analysis of completion rates across different device types revealed significant patterns. Desktop users consistently showed higher completion rates compared to mobile users across most surveys. Statistical testing

confirmed these differences were significant in all surveys except the most recent Long-distance travel study, which used a panel with financial incentives.

While these differences cannot be attributed solely to device effects without controlling for demographic factors, the paradata provides valuable direction for interface optimization to improve completion rates for all devices.

4.2 Questionnaire Design Improvements

Paradata analysis enabled direct comparison of different question formulations. For example, in collecting occupation data, the Quebec City survey used a traditional single question with multiple options, while the Montreal survey split the question into a sequence of shorter questions.

For the two formulations, it is possible to compare the total response time for the occupation questions as well as the answer modification rates, and thus give an idea on the cognitive burden. Table 2 includes examples of those metrics for different age/gender groups, broken down by occupation. The dominant occupation category tends to experience the least burden in each group. However, in the 'other' category, more than 20% of respondents changed their mind, suggesting potential confusion and areas for improvement. The population of both surveys are not comparable and thus this data does not allow statistical significance, but with A/B testing, this type of analysis allows assessment of question impacts and refinement of future questionnaire design.

Table 2: Example of paradata statistical analysis of the occupation questions, for selected respondent age/gender groups

Category	Qc23				Mtl23			
	% of group	answer time		% changes	% of group	answer time		% changes
		med	avg			med	avg	
Occupation								
Women aged 65+	N=6549				N=31 448			
- retired	89.1	4	4.9	1.14	83.9	12	13.96	2.59
- worker (part or full time)	7.8	7	7.88	7.22	10.7	9	10.98	4.79
- other	3.1	8	10.33	20.81	5.4	17	21.59	20.17
Women between 25 and 44	N=8376				N=40 227			
- worker (part or full time)	90.1	3	3.9	2.27	77.2	5	6.23	4.81
- student (part or full time)	4.4	5	7.23	22.51	4.3	9	11.82	15.88
- worker AND student	-	-	-	-	9.1	8	10.12	13.67
- other	5.5	8	12.04	26.5	9.4	13	16.42	20.17

This type of empirical, data-driven questionnaire refinement demonstrates how paradata can enhance survey design beyond traditional testing methods.

5 Discussion and Conclusion

The Evolution platform represents a significant advancement in transportation survey methodology by addressing key challenges in data quality and respondent experience. Its open-source nature ensures transparency and enables community-driven enhancements, while its comprehensive feature set overcomes limitations found in both commercial and existing open-source survey tools.

The platform's paradata collection capabilities provide unprecedented insight into the survey process itself, enabling continuous improvement based on empirical evidence rather than assumptions. By analyzing response patterns, device effects, and question performance, researchers can iteratively enhance both the survey instrument and the resulting data quality.

Future directions for the platform include increased A/B testing of question formulations, enhanced mobile interface optimization, and integration with passive data collection methods. Privacy-preserving approaches to location data collection also represent an important area for development, particularly for jurisdictions with stricter data protection regulations.

By making robust travel survey methodologies more accessible to the broader research community, Evolution contributes not only to improved data collection, but also to methodological transparency and knowledge sharing within transportation research.

Acknowledgement

The authors wish to acknowledge the support of various survey partners: MTMD (Quebec Ministry of Transportation and Sustainable Mobility), ARTM (Montreal Region Transportation Authority), and all other survey partners over the years for providing valuable insights and feedback.

References

Armoogum, J., Wittwer, R., Madre, J.-L., and Gopal, Y. (2024). Sampling, nonresponse, and quality in travel surveys. *Transportation Research Procedia*, 76:638–643. 12th International Conference on Transport Survey Methods.

Barthelmes, L., Kübler, J., Bönisch, L., Kagerbauer, M., and Vortisch, P. (2024). An open-source interactive travel diary for web-based trip reporting. *Transportation Research Procedia*, 76:373–384. 12th International Conference on Transport Survey Methods.

Bayart, C. and Bonnel, P. (2015). How to combine survey media (web, telephone, face-to-face): Lyon and rhône-alps case study. *Transportation Research Procedia*, 11:118–135. Transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods 16-21 November 2014, Leura, Australia.

Bayart, C. and Bonnel, P. (2024). Mixed-mode household travel surveys: synthesis from three experiments in france. *Transportation Research Procedia*, 76:545–554. 12th International Conference on Transport Survey Methods.

Chung, B., Srikukenthiran, S., Habib, K., and Miller, E. (2017). The development of a web-survey builder (staisi): designing household travel surveys for data accuracy and reduced response burden. In *11th International Conference on Transport Survey Methods*.

Colaço, R. and de Abreu e Silva, J. (2024). Intrapersonal variability and underreporting – comparing a one-week shopping survey with a one-day travel survey. *Transportation Research Procedia*, 76:133–142. 12th International Conference on Transport Survey Methods.

Evolution (2025). The evolution platform. `https://github.com/chairemobilite/evolution`. version 0.5.0.

Keshavarz Mohammadian, H. (2023). Definition and analysis of respondent burden in web-based travel surveys. Master's thesis, Polytechnique Montréal.

Vardhan, H., Rai, I., Kathait, N., and Agarwal, A. (2022). Crowd-sourced web survey for household travel diaries. *Transportation Research Procedia*, 62:245–252. 24th Euro Working Group on Transportation Meeting.

Vehovar, V., Berzelak, N., and Čehovin, G. (2024). Identifying a set of key paradata indicators in web surveys based on the relationship with response quality, respondent characteristics and survey estimates. *Survey Research Methods*, 18(3):225–249.

Verzosa, N., Greaves, S., Ho, C., and Davis, M. (2021). Stated willingness to participate in travel surveys: a cross-country and cross-methods comparison. *Transportation*, 48(3):1311–1327.