

ALTO HSR CITIZEN RESEARCH INITIATIVE

An independent, non-partisan citizen research project

APPENDIX — DATA INTEGRITY

Detection of a Coordinated Submission Attempt

Forensic analysis of the ALTO HSR Participant Experience Survey (n=386) and the methodology by which 32 responses were analytically excluded from the published figures

COMPANION TO

Participant Experience Survey — Full Analysis

citizenresearch.ca/consultation/participant-experience-survey

Published April 17, 2026

ALTO public consultation deadline: April 24, 2026

Summary

In the final days of the ALTO public consultation period, the Participant Experience Survey conducted independently by the ALTO HSR Citizen Research Initiative received a distinct batch of responses whose behavioural signature is inconsistent with organic participation. This appendix documents the detection, characterization, and analytical handling of those responses. The underlying dataset has been preserved intact; filtering is performed analytically and is fully reversible.

HEADLINE FINDING

Of 386 responses collected between March 24 and April 17, 2026, 32 meet a three-criterion objective signature of inauthentic submission: (a) arrival on or after April 11, 2026, (b) total completion time of 120 seconds or less on the 17-question survey, and (c) a uniformly positive rating profile on Q4 (information adequacy), Q6 (map usefulness), and Q13 (process legitimacy). Zero critical respondents and zero pre-April 11 respondents of any sentiment meet this signature. The 32 flagged responses have been analytically excluded from the published figures; the retained analytical sample is n=354.

Why this matters

A consultation process whose participation mechanisms are subject to manipulation from more than one direction cannot produce a reliable record of public opinion. This appendix documents such manipulation directed at an independent citizen research survey. It sits alongside the “Strongly Disagree” suppression finding in ALTO’s own online survey — where 36.0% of completers experienced a truncated five-point Likert scale — as evidence that the participation infrastructure of this consultation has not been neutral.

What this appendix does not claim

The CRI does not attribute the coordinated submission effort to any specific individual or organization. The behavioural signature is consistent with a small group of real people following shared guidance, rather than automated bot activity — but the analysis does not rest on any claim about who was responsible, only on evidence that the pattern exists and is distinct from the rest of the dataset.

1. Detection — The Step Change

Two metrics, considered together, produce an unambiguous inflection on April 11, 2026. Neither metric in isolation would be decisive; together they are.

1.1 Sentiment shift

Each response was classified on the three structured sentiment questions — Q4 (information adequacy), Q6 (map usefulness), and Q13 (process legitimacy) — using the following rule: a respondent rating ALTO positively (option 1 or 2) on at least two of the three, with no negative rating on any, is classified as *pro-ALTO*. A respondent rating ALTO negatively on at least two, with no positive rating on any, is classified as *critical*. Everyone else is classified as *mixed or neutral*.

Applied to the full 386-respondent dataset, the daily pro-ALTO share was at or below 17% every day from March 24 through April 10. On April 11 onward the ratio inverted.

Period	n	Pro-ALTO	Pro-ALTO share	Critical
March 24 – April 10	200	8	4.0%	143
April 11 – April 17	186	48	25.8%	105

The shift is a 6.4× increase in pro-ALTO share across periods. A 2×2 chi-square test of independence returns $\chi^2 = 36.95$ on 1 degree of freedom — substantially beyond the critical value of 10.83 at $p = 0.001$. The probability that a shift of this magnitude arises from random sampling variation is below one in a thousand.

1.2 Completion time — the hardest evidence

A 17-question survey containing open-ended text fields requires meaningful time to complete authentically. Across all 248 critical respondents over the full survey window, **zero completed the survey in under 120 seconds**. The same holds across the 82 mixed-or-neutral respondents: the minimum completion time in either group is over two minutes. Across the 8 pro-ALTO respondents submitted before April 11, the median completion time was 510 seconds — consistent with the rest of the dataset.

Across the 48 pro-ALTO respondents submitted on or after April 11, the median completion time is **81 seconds**. One-third (16 of 48) completed in 60 seconds or less. The fastest completed in 28 seconds.

Group	n	Median	10th percentile	90th percentile	Minimum
Critical (full period)	248	556 s	286 s	2,416 s	122 s
Mixed / neutral (full period)	82	463 s	268 s	1,061 s	121 s

Group	n	Median	10th percentile	90th percentile	Minimum
Pro-ALTO, before Apr 11	8	510 s	—	—	204 s
Pro-ALTO, Apr 11 onward	48	81 s	40 s	520 s	28 s

KEY INTERPRETATION

If the 48 post-April 11 pro-ALTO responses were a genuine late-consultation surge of supportive opinion, they would complete the survey at approximately the same speed as the rest of the dataset. Authentic late-consultation respondents would still need to read questions, consider answers, and submit text. The 90-second median indicates that these respondents are not engaging with the questions in any substantive way. This is the single hardest metric in the analysis — it is not a matter of sentiment interpretation. Completion time is either high or it is not.

2. Characterization — Six Supporting Signals

Once the post-April 11 pro-ALTO cohort was isolated, six further attributes were examined. All six reinforce the conclusion that this cohort is distinct from the rest of the dataset in ways that are difficult to produce organically.

2.1 Open-ended text engagement

Of the 48 post-April 11 pro-ALTO respondents: 23 left Q16 (the single most significant gap or failure) completely blank; 33 submitted 5 characters or fewer (“X,” “—,” “None”); 38 submitted 15 characters or fewer. By contrast, the 8 pre-April 11 pro-ALTO respondents wrote substantial, specific responses — including one explicitly critical of the Citizen Research Initiative itself (“this group has a political agenda while pretending to be non-partisan”). The pre-April 11 pro-ALTO respondents read as authentic minority-view participants. The post-April 11 cohort does not.

2.2 Boilerplate phrase fingerprint

Distinctive phrases are dramatically over-represented in the post-April 11 pro-ALTO text pool relative to the critical pool.

Phrase fragment	Pro-ALTO occurrence	Critical occurrence	Over-representation
“early” (stage, phase, days, in the process)	9 / 56 (16.1%)	3 / 248 (1.2%)	13×
“great”	7 / 56 (12.5%)	3 / 248 (1.2%)	10×
“kiosks”	2 / 56 (3.6%)	0 / 248 (0%)	—
“first wave” / “first round”	2 / 56 (3.6%)	0 / 248 (0%)	—
“looking forward”	2 / 56 (3.6%)	0 / 248 (0%)	—
“comprehensive”	1 / 56 (1.8%)	0 / 248 (0%)	—

The phrase fragments over-represented in the pro-ALTO cohort are a near-exact mirror of ALTO’s own framing of the consultation (“early in the process,” “first wave,” “comprehensive”). This is consistent with respondents following talking points rather than answering open-ended questions in their own words.

2.3 Notification source inversion

Q1 asks respondents how they first learned about the ALTO consultation. Across the full dataset, 22 of 383 respondents (5.7%) selected “Direct notification from ALTO (email, mail, or phone).” Of those 22, 21 are pro-ALTO. Only 1 of 245 critical respondents (0.4%) reports having been directly notified by ALTO. After the forensic filter is applied, the overall direct-notification rate drops to 2.0% — consistent with the figure independently derived from the April 13 survey analysis (2.4%) and with the overwhelming pattern documented in Q1 open-ended responses,

where respondents describe learning about the consultation through neighbours, farmers' networks, MPP posts, and community Facebook groups.

2.4 Geographic distribution

The critical cohort reports living in or adjacent to the study corridor at a rate of 76.4%, with a further 12.2% within 5 km. Only 0.8% report living outside Ontario. In the pro-ALTO cohort, 14.3% report living outside Ontario — a 17× over-representation. A cohort motivated by substantive concerns about route impact on their properties is unlikely to be concentrated in jurisdictions outside the corridor's footprint.

2.5 Tight temporal clusters

Where responses arrive through organic, independent participation, the gaps between successive responses follow a Poisson-like distribution. In the post-April 11 pro-ALTO cohort, responses arrive in clustered bursts inconsistent with independent participation.

Date	Window	Pro-ALTO responses	Span	Mean interval
April 12	14:29–14:52	4	23.4 min	~7 min
April 12	22:09–22:24	4	15.1 min	~5 min
April 14	09:36–09:58	3	22.1 min	~11 min (all ≤56 s, all Q15 blank)
April 14	13:12–13:36	4	23.5 min	~8 min
April 17	18:28–18:33	5	4.3 min	~1 min (fastest burst)
April 17	19:19–19:23	3	3.7 min	~2 min

2.6 Pre-April 11 pro-ALTO responses are authentic

The 8 pro-ALTO responses received before April 11 differ from the post-April 11 cohort on every dimension analyzed: completion time, text length, phrase fingerprint, notification source, and geographic distribution. Each of the 8 wrote substantive, specific, internally coherent responses. One, dated March 30, took 45 minutes to complete. Another, dated March 31, wrote a sustained critique of the CRI's methodology. These 8 responses have been retained in the analytical sample; they read as authentic minority-view participants, and excluding them would suppress legitimate dissent.

3. Analytical Filter — Objective, Replicable, Conservative

The analytical filter used to derive the published headline figures is stated in full below. It is fully specified by observable attributes; no subjective judgment is applied to any individual response.

THE THREE-CRITERION RULE

A response is excluded from the analytical sample if and only if ALL THREE of the following are true:

- (a) **submission date (survey_start_time) is on or after April 11, 2026; AND**
- (b) **total completion time (end_time – start_time) is 120 seconds or less; AND**
- (c) **the respondent is classified as pro-ALTO on the Q4/Q6/Q13 sentiment trilogy — i.e., rated option 1 or 2 on at least two of the three questions, with no option 3 or 4 rating on any of the three.**

Under this rule, 32 of 386 responses are excluded. The analytical sample is n=354.

Specificity

The rule catches no critical respondents. It catches no mixed-or-neutral respondents. It catches no pro-ALTO respondents who arrived before April 11. It catches 16 fewer post-April 11 pro-ALTO respondents than a pure date-plus-sentiment rule would — specifically, pro-ALTO respondents who took longer than 120 seconds to complete the survey are retained even if they arrived after April 11. The rule is deliberately more conservative than the evidence would permit.

Reversibility

The raw dataset (n=386) has been preserved intact. It has not been edited, deleted, or altered in any way. Every analysis in this document, and every figure on the Participant Experience Survey webpage, can be reconstructed from the raw dataset and the filter rule by any analyst with access to both. The filter is implemented as a single conditional in Python/pandas; the code is ~5 lines.

What the filter does not do

The filter does not attempt to classify individual motivation. It does not attempt to identify which real-world person or entity authored any specific response. It does not attempt to distinguish between responses produced by a single author submitting multiple times and responses produced by multiple authors following shared guidance. All of those questions are outside the scope of this analysis, which addresses only the narrower question of whether certain responses are analytically distinguishable as a cohort and whether including them would distort the published figures.

4. Effect on the Published Figures

The table below shows the effect of the forensic filter on each of the headline figures published in the Participant Experience Survey analysis. The directional effect is consistent across all six: the filtered figures are uniformly stronger than the as-collected figures. This is expected — a coordinated pro-ALTO submission attempt depresses critical-sentiment percentages in direct proportion to its scale.

Indicator	As collected (n=386)	After filter (n=354)	Direction
Q1 — Heard directly from ALTO	5.7%	2.0%	↓
Q4 — Information inadequate / very inadequate	80.5%	87.8%	↑
Q6 — Map not very useful / not useful at all	66.7%	72.7%	↑
Q13 — Do not believe consultation is genuine	78.5%	85.7%	↑
Q13 — Believe process not designed for opposition	41.6%	45.4%	↑
Q3 — Saw only four of five Likert options (completers)	32.4%	36.0%	↑

The filtered figures also align more closely with the results independently reported in the April 13, 2026 analysis (n=332), which predated most of the coordinated submission effort. The April 13 report found 86.4% rating information inadequate, 66.4% rating the map not useful, and 84.8% not believing the consultation was designed to register community input. The filtered figures in this appendix — 87.8%, 72.7%, and 85.7% respectively — sit within a few percentage points of the April 13 baseline, which is itself a form of external validation for the filter.

A NOTE ON DISCLOSURE PHILOSOPHY

Some research projects would treat the discovery of a coordinated submission attempt as embarrassing and bury it. The CRI takes the opposite view. The fact that this survey was targeted — and that the targeting was detectable, bounded, and analytically excludable — is itself evidence about the broader consultation environment. Both the raw figures and the filtered figures are published. The underlying dataset and the filter script are available for independent verification. A consultation process whose integrity depends on no one looking is not a consultation; it is a formality. The same standard applies to an independent citizen research survey.

5. Reproducibility and Data Access

All analyses in this appendix are reproducible from the raw SurveyMonkey export. The CRI will provide the following to any qualified researcher, journalist, or oversight body on request:

1. The raw dataset (SurveyMonkey standard export, n=386, timestamps and response content; personal identifiers where provided by respondents have been removed).
2. The Python analysis scripts used to classify sentiment, compute completion times, identify temporal clusters, and apply the forensic filter.
3. The full list of 32 flagged respondent IDs, with the specific criterion each one satisfies.
4. A reproducibility notebook regenerating every figure and statistic in this appendix from the raw dataset.

Requests

Requests for access should be sent to the CRI via citizenresearch.ca. Any redaction of personal identifiers will be documented. No individual respondent will be identified in any publication or disclosure without their explicit written consent.

Limitations

SurveyMonkey's standard export does not include IP addresses in the variant used here, and the CRI did not collect respondent email addresses as a required field. Analysis is therefore based on timestamps, completion durations, sentiment classifications, and response content alone. Access to IP-level data would permit additional clustering analysis (e.g., testing whether flagged responses originated from a small number of network locations), which could provide additional characterization. The absence of that layer does not affect the conclusions of this appendix, which rest on behavioural metrics that are visible without it.

ALTO HSR CITIZEN RESEARCH INITIATIVE

An independent, non-partisan citizen research project examining the ALTO HSR proposal in the public interest.

citizenresearch.ca • altohsrcitizenresearch.ca