Memorandum

To: Josh Roll, Oregon Department of Transportation

From: Patrick Singleton, Utah State University

Date: 11 December 2025

Re: Anomalies in Aggregated Pedestrian Volume Estimates from Push-Button Data

---

Table of contents

Background

As part of Oregon Department of Transportation (ODOT) research project SPR 857, "Active Transportation Counts from Existing On-street Signal and Detection Infrastructure," Utah State University (USU) was retained to provide technical assistance for the implementation of the research. Specifically, the scope of work involved consulting on potential anomalies in the pedestrian signal data and/or volume estimates.

Working with ODOT, Portland State University (PSU) implemented the research and created a pedestrian push-button dashboard on the BikePed Portal website (https://bikeped.trec.pdx.edu/pushbutton/). The system ingests data from hundreds of traffic signals throughout Oregon, processes push-button data, and applies volume estimation equations (developed during the research project). The results are accessed through a dashboard that allows users to select locations and visualize hourly or daily estimated pedestrian volumes at each location, overall and by phase. Figure 1 shows an example of the dashboard interface.

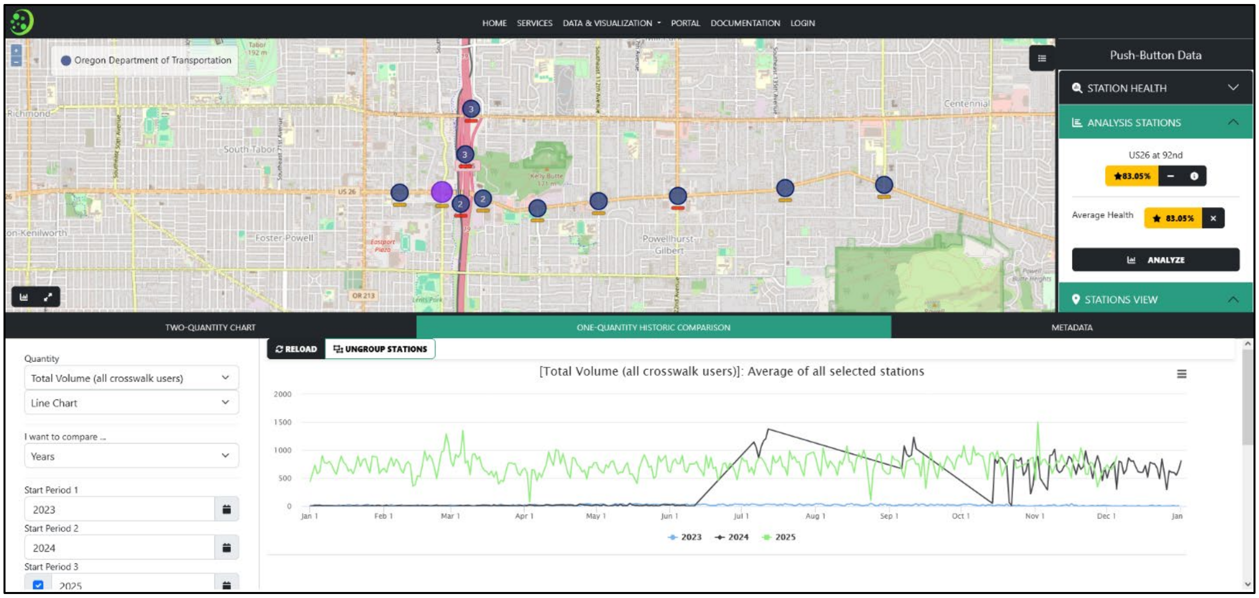


Figure 1: View of the PSU BikePed Portal's pedestrian push-button dashboard, showing an example of an anomalous trend at one location.

Recently, USU was informed of some anomalies in the estimated pedestrian volumes being generated and visualized in the push-button dashboard. An example of this is shown in Figure 1. At this location (SE Powell Blvd and SE 92nd Ave, Portland), the average daily pedestrian volume is relatively low throughout 2023 and the first half of 2024. Following a data gap of several months, the pedestrian volume since late 2024 is significantly larger, by several orders of magnitude. This is unexpected, assuming no major changes to the signalized intersection (e.g., its operation, its surroundings, or construction), and due to the automation of data ingestion and estimation processes.

USU was asked to investigate this and similar anomalies in the estimated pedestrian volumes being presented by the push-button dashboard. We conducted this work in two parts. First, we took initial broad steps to identify the scale, scope, and characteristics of these anomalies. Second, we did deeper investigations to explore potential explanations for the identified anomalies.

This memo documents the methods employed, results found, and recommendations for dealing with or further diagnosing potential anomalies in pedestrian push-button data in Oregon.

Part I: Scope of the issue

Methods

Following conversations with ODOT staff, we focused our attention on identifying the scale of the anomalies and characterizing them. As a first pass, we looked just at the outputs—aggregated estimated pedestrian volumes—since that is what end users would be interested it.

After some spot checks, our attention quickly turned to investigating changes following a period (roughly June to October 2024) during which ODOT's traffic signal data were switched to a new system. At several locations, estimated pedestrian volumes were different, often (much) higher afterwards. (See data gap and scale change in Figure 1.) Staff did not expect any change in the data being ingested and processed, nor changes to the signals themselves, so this is considered anomalous.

To investigate more systematically, we worked with PSU's BikePed Portal staff to obtain credentials to access the push-button database underlying the dashboard. After obtaining access, we explored available data. There are four main tables utilized:

- *intersections*: Unique traffic signals with data. Columns include: unique IDs, name/description, signal group (corridor or zone), and latitude/longitude.
- *has_data*: Indication of whether data are available. Counts the number of records available at each intersection (device_id) for a given (start) date-time and interval (time resolution, e.g., 15 minutes).
- *ped_data_raw*: Push-button data ingested from ODOT. For each intersection (device_id), crossing (phase), and (start) date-time (in 15-minute intervals), provides the number of pedestrian services, pedestrian actuations, and unique pedestrian actuations.
- *ped_data_estimated*: Estimated pedestrian volumes processed. For each intersection (device_id), crossing (phase), (start) date-time, and interval (time resolution), provides the estimated pedestrian volume (total and different definitions of "pedestrian"). The estimates are from the number of unique pedestrian actuations, calculated using the estimation equation (from the research project). Time intervals include hours, days, and months. A count of data availability is also provided, presumably based on the *has_data* table.
- Note about time: In the tables, the date-time column appears to be in UTC format. However, intervals that are days appear to be based on local Oregon time (most Pacific, some Mountain).

Our investigative approach involved checking for major changes in estimated pedestrian volume values and trends before versus after the mid-2024 data system change. To support this, we identified a one-year period both before and after the change for comparative purposes. We calculated the number of intersections with no daily estimated pedestrian volumes for each day, from January 2023 through November 2025. Figure 2 shows the resulting plot. The data outage lasted from roughly 6 June through 12 October. Thus, we selected the following dates for each period:

- *Before*: 1 June 2023 through 31 May 2024.
- *After*: 1 November 2024 through 31 October 2025.

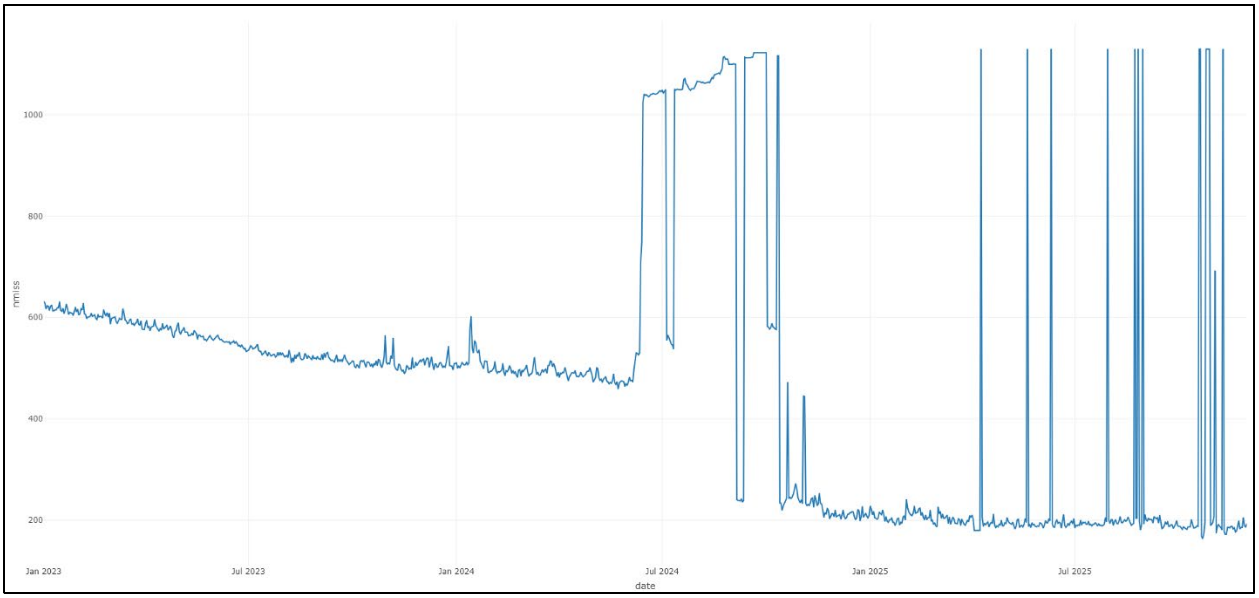


Figure 2: Number of intersections with missing data between January 2023 and November 2025, used to identify before and after periods for comparison.

After identifying the before and after periods for comparison, we then used SQL queries to assemble total (all phase) estimated pedestrian volumes for each intersection on each day in each period. Next, we calculated summary statistics to describe and compare the trends and patterns, for comparison. Finally, we analyzed the results, including summarizing the statistics and their differences, and grouping intersections with similar differences. This process is detailed in the following steps:

- From the *ped_data_estimated* table, filter for 1-day time intervals within the before/after periods.
- Sum estimated volumes (total_estimated_volume) and data availability (sample_count) across all phases, for each intersection and day.
- For each intersection and period (before = 1, after = 2), calculate the mean (M), standard deviation (SD), coefficient of variation (CV) (= SD ÷ mean), and number of days (N) with data.
- For each intersection, calculate differences between the before and after periods:
    - Mean: difference (Mdif = M1 − M2), ratio (Mrat = M1 ÷ M2), ratio of the difference (Mdifrat = Mdif ÷ M2), and p-value (Mp) from a two-sample (independent samples) t-test of the equality of means.
    - Coefficient of variation: difference (CVdif = CV1 − CV2), ratio (CVrat = CV1 ÷ CV2), ratio of the difference (CVdifrat = CVdif ÷ CV2), and p-value (CVp) from an asymptotic (Feltz-Miller) test of the equality of coefficients of variation.
- Classify intersections into categories based on comparative statistics.
- Select representative intersections in each category, for more detailed investigation.

Results

There were 1,129 signalized intersections listed in the push-button database. Many did not have data during the before period, and some had too many missing days for a reliable comparison.

Thus, we filtered the locations to be the 585 intersections with at least 8 months (244 days) of complete data in both before and after periods.

Table 1 reports descriptive statistics about the summary and comparison statistics for the pedestrian volume trends at the 585 intersections. Some notable findings:

- Average daily estimated pedestrian volumes (means):
    - Most intersections (71%) had lower volumes in the before period (mean differences: mean -27, median -37).
    - The median intersection had 24% lower volumes in the before period, and 30% of intersections had before volumes that were less than half of after volumes.
    - There was a statistically significant ($p < 0.05$) difference in average volumes (before vs. after) at 92% of intersections.
- Variability in daily estimated pedestrian volumes (coefficients of variation):
    - Most intersections (85%) had more volume variation in the before period (CV differences: mean 0.20, median 0.14).
    - The median intersection had 71% more volume variation in the before period, and 37% of intersections had volume variation in the before period that was more than twice that in the after period.
    - There was a statistically significant ($p < 0.05$) difference in volume variation (before vs. after) at 90% of intersections.

Table 1: Descriptive statistics of the summary and comparison statistics (N = 585 intersections)

Statistic	*Mean*	*Min.*	*25th %-ile*	*Median*	*75th %-ile*	*Max.*
M1	184.16	5.72	62.65	121.05	205.91	4679.13
M2	211.26	27.60	107.51	163.68	237.24	3962.83
SD1	93.82	3.08	23.04	38.32	74.92	9155.93
SD2	67.96	2.04	21.25	30.21	60.13	5305.85
CV1	0.46	0.15	0.27	0.36	0.50	6.66
CV2	0.26	0.05	0.16	0.20	0.30	3.82
N1	355.65	245.00	360.00	366.00	366.00	366.00
N2	346.52	244.00	350.00	351.00	352.00	352.00
Mdif	-27.10	-877.47	-78.63	-37.49	10.58	1595.31
Mrat	1.05	0.02	0.44	0.76	1.06	28.51
Mdifrat	0.05	-0.98	-0.56	-0.24	0.06	27.51
Mp	0.03	0.00	0.00	0.00	0.00	0.93
CVdif	0.20	-3.62	0.05	0.14	0.27	6.26
CVrat	2.36	0.05	1.20	1.71	2.49	54.48
CVdifrat	1.36	-0.95	0.20	0.71	1.49	53.48
CVp	0.04	0.00	0.00	0.00	0.00	1.00

For the classification or grouping of intersections with similar patterns, we tried several different approaches. First, we attempted clustering, an unsupervised machine learning method that groups observations based on similar patterns in their data. Specifically, we tried both k-means clustering and hierarchical clustering, using different subsets of the comparative statistics shown in Table 1, and trying different clustering methods. Although potentially interesting, the cluster analysis results were quite varied and tended to produce one large cluster and multiple very small or single-observation clusters.

Instead, we opted to use a heuristic approach for classifying intersections, based on one mean statistic (mean ratio, Mrat) and one CV statistic (CV difference, CVdif). This allowed us to group observations with different absolute volume differences but similar relative volume or relative variation differences. On each of these two dimensions, we selected two threshold values to yield three categories, comparing before to after: lower, similar, and higher.

Figure 3 and Table 2 present the results of this deterministic categorization. Regarding the mean ratio, the categories confirm that many intersections (49%) had much lower volumes before than after, while some (16%) had much higher volumes before than after. Most intersections (79%) had roughly similar coefficients of variation, but having more variation in the before period was slightly more common, especially for locations with lower before volumes. Only about 30% of intersections had similar means and similar coefficients of variations before and after.

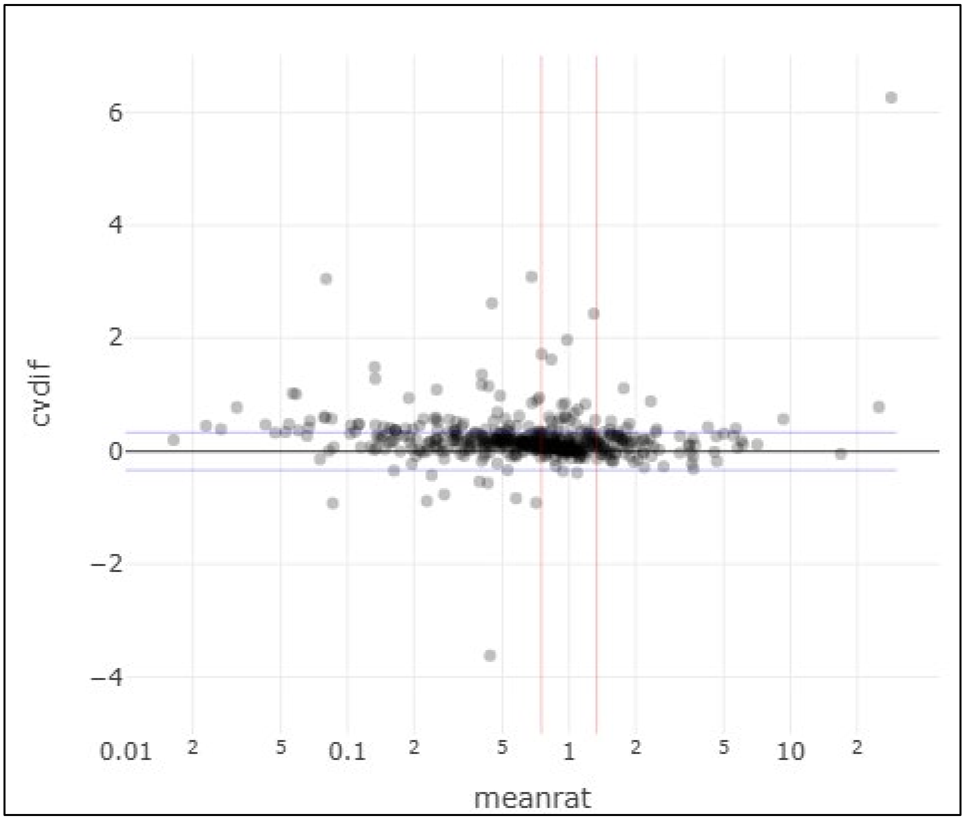


Figure 3: Plot of mean ratio (Mrat) versus CV difference (CVdif) for intersections, with threshold values shown

Table 2: Table of mean ratio (Mrat) versus CV difference (CVdif) categories for intersections

Number (%) of intersections *(values are comparing before to after)*		*Mean ratio (Mrat)*		
		Lower (0.00 to 0.75)	*Similar (0.75 to 1.33)*	*Higher (1.33 to Inf)*
Coefficient of variation difference (CVdif)	*Lower (−Inf to −0.33)*	11 (1.9%)	2 (0.3%)	0 (0.0%)
	Similar (−0.33 to +0.33)	206 (35.2%)	178 (30.4%)	79 (13.5%)
	Higher (+0.33 to +Inf)	69 (11.8%)	25 (4.3%)	15 (2.6%)

We then reviewed the intersections in each “anomaly” category, and selected a subset to potentially investigate in more detail. The tentatively selected intersections in each category are:

- Mean lower, CV lower
    - Yew Ave / Airport Wy and US-97 NB Ramps, Redmond (10097)
    - SW Baseline St (OR-8 WB) and S 1st Ave, Hillsboro (2B404)
- Mean lower, CV similar
    - Delta Hwy (OR-132) and Green Acres Rd, Eugene (05166)
    - OR-99E and 10th St, Oregon City (2B015)
    - OR-224 and SE Monroe St, Milwaukee (2B050)

  - Pacific Hwy W (OR-99W) and St Main St and SW Johnson St, Tigard (2B349)
  - State St (OR-43) and Foothills Rd, Lake Oswego (2B485)
- Mean lower, CV higher
  - Altamont Dr and Laverne Ave, Klamath Falls (11039)
  - Adams Ave (US-30) and Fir St, La Grande (13004)
  - SE Powell Blvd (US-26) and SE 92nd Ave, Portland (2B100)
- Mean similar, CV lower
  - S Hwy 101 (US-101) and SE/SW 51st, Lincoln City (04092)
  - Court Ave (US-30 WB) and Main St, Pendleton (12008)
- Mean similar, CV similar
  - No intersections selected, assumed to not be anomalous.
- Mean similar, CV higher
  - E 2nd St (US-30 WB) and Washington St, The Dalles (09002)
  - Main St (US-30, OR-7) and Auburn Ave, Baker City (13018)
- Mean higher, CV lower
  - No intersections identified.
- Mean higher, CV similar
  - US-30 and Columbia Blvd, St Helens (01069)
  - US-199 and OR-99 SB, Grants Pass (08021)
  - SW 4th Ave and SW 9th St and S Park Blvd, Ontario (14005)
- Mean higher, CV higher
  - N College St (OR-219) and E Mountainview Dr, Newberg (03093)
  - Pacific Blvd S (OR-99E) and Allen Ln, Albany (04070)

Example time series plots of one of each type of intersection are shown in Figure 4 (a-g).

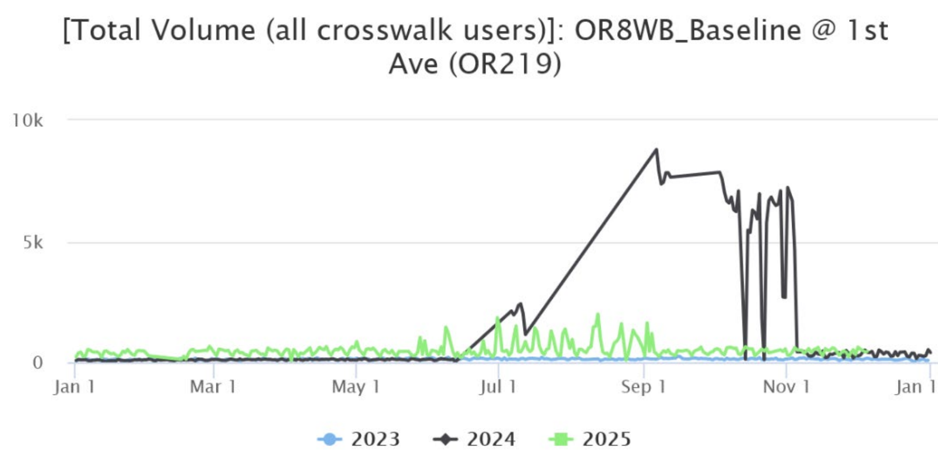


Figure 4a: Estimated pedestrian volumes for a "mean lower, CV lower" intersection (2B404)

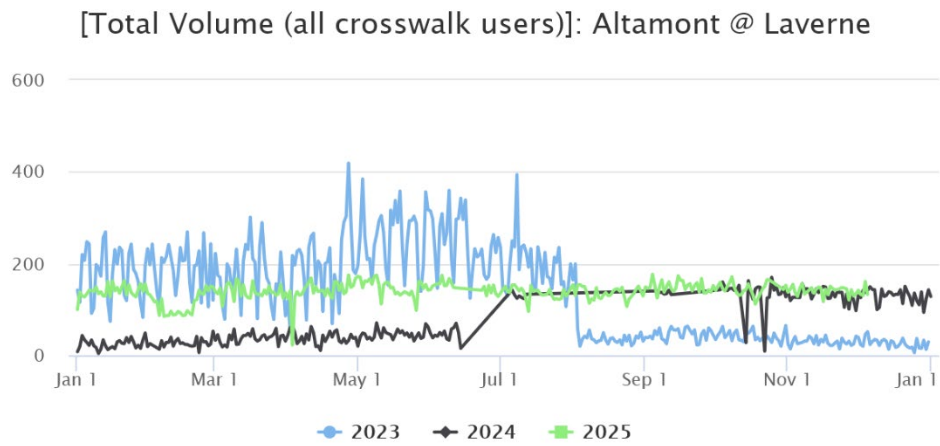


Figure 4b: Estimated pedestrian volumes for a "mean lower, CV similar" intersection (11039)

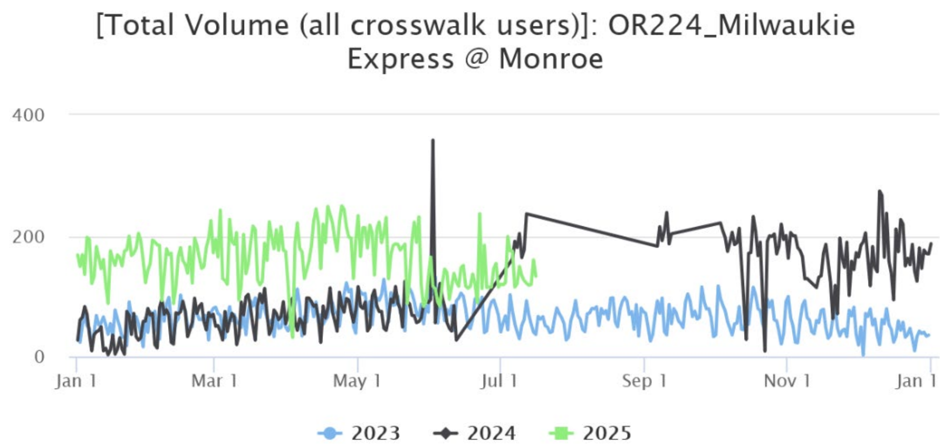


Figure 4c: Estimated pedestrian volumes for a "mean lower, CV higher" intersection (2B050)

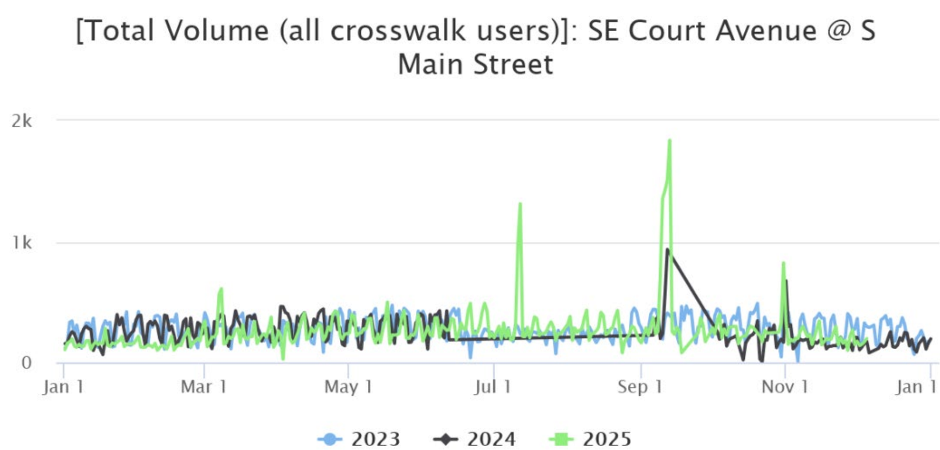


Figure 4d: Estimated pedestrian volumes for a "mean similar, CV lower" intersection (12008)

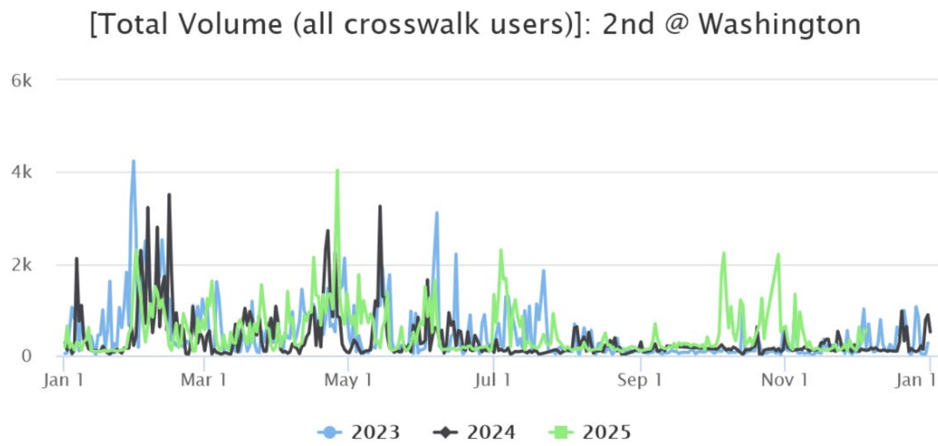


Figure 4e: Estimated pedestrian volumes for a "mean similar, CV higher" intersection (09002)

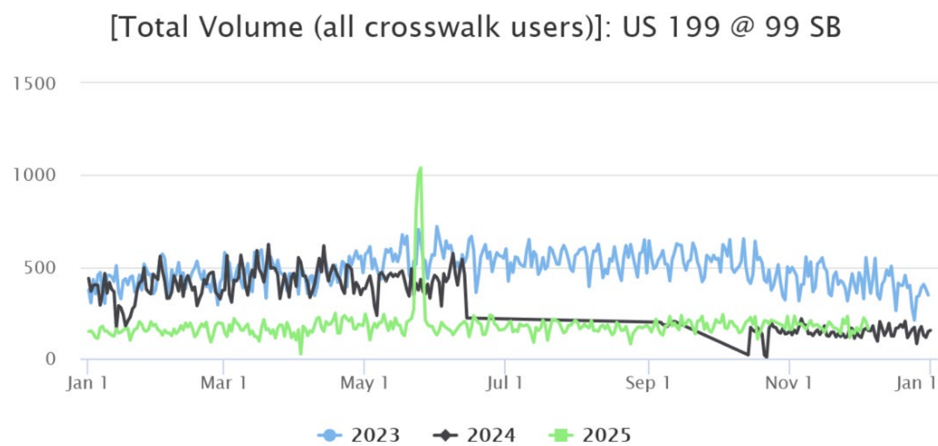


Figure 4f: Estimated pedestrian volumes for a "mean higher, CV similar" intersection (08021)

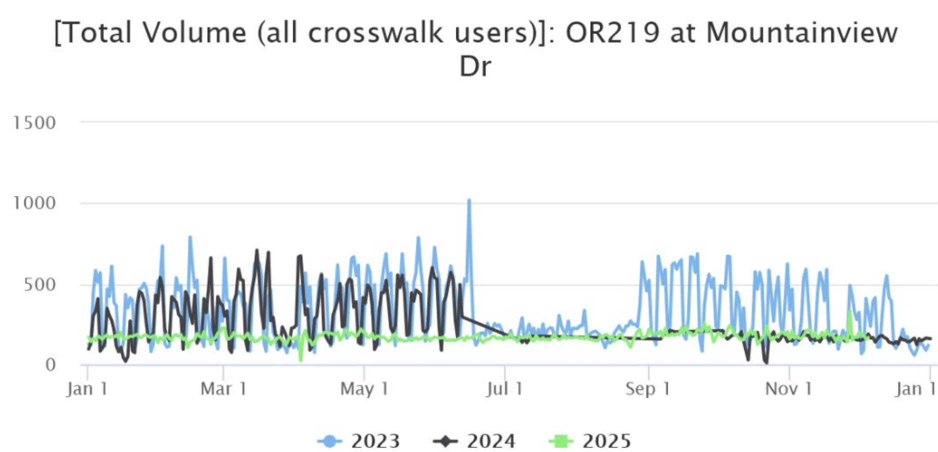


Figure 4g: Estimated pedestrian volumes for a "mean higher, CV higher" intersection (03093)

Part II: Investigation of potential explanations

Potential explanations

Before proceeding with detailed investigations of data at the selected example locations, we brainstormed a list of potential explanations for why we might see big differences in the magnitude and/or variation of estimated pedestrian volumes in the push-button dashboard. This list of potential explanations helped us to identify and think through strategies to diagnose this issue, including ways to find evidence consistent or inconsistent with each hypothesis. These broad types of potential explanations include:

- *Real changes in behavior or operations*: It is possible that there have been real changes in pedestrian volumes, push-button behavior, and or traffic signal operations that are showing up in the trends analyzed in Part I. This might result from, for example: new land uses, re-aligned transit routes or stops, adjusted signal phasing or timing, etc.
  - This is not possible to diagnose through Portal data alone. Ground-truth pedestrian volumes would need to be collected and compared to the traffic signal data. However, we do have access to some ground-truth pedestrian volumes for two days from 30+ intersections in 2022, which could be compared to pedestrian volume estimates for similar time periods.
- *Data processing before Portal*: It is possible that some step in the data processing before the data is sent to Portal is not working as expected. This could include steps in the translation of traffic signal controller logs into the databases that are ingested by Portal, or a misalignments of signal locations or IDs.
  - This is difficult to diagnose through Portal data alone. One might want to start with the raw traffic signal controller log data and follow it through the data processing pipelines. However, we do have access to some raw traffic signal data for two days from 30+ intersections in 2022, which could be compared to the raw data in Portal for similar time periods.
- *Data processing within Portal*: It is possible that some step in the data processing within Portal is not working as expected. This could include the temporal aggregation process or the application of the pedestrian volume estimation equations.
  - This could be diagnosed through spot checks of the data construction process, from the raw data to the daily estimates. Data processing scripts may also have to be consulted.
- *Data completeness*: Incompleteness of data is somewhat common in datasets such as these, arising from traffic signal controller data logs. Missing data might arise for a variety of reasons: a broken or malfunctioning push-button, construction, loss of connectivity, etc. Various assumptions about what and when data are missing could affect the outputs of this process. Systematic or long-term missing data (at the phase or intersection level) could result in very different pedestrian volume estimates for different periods. These issues could arise from data processing before or within Portal.
  - Some of this may be able to be diagnosed using Portal data. Within Portal, this could be diagnosed by identifying missingness at the most granular level (e.g., 15-minutes by phase), and comparing that to the differences in aggregate pedestrian volumes that result. Additionally, the treatment of missingness within Portal could be traced using spot checks from the raw data to the daily estimates. Before

Portal, information about how the *has_data* table is constructed could help to diagnose the treatment of missing data.

There may other potential explanations for the detected anomalies that are not described here. We welcome suggestions about other potential explanations.

Methods

Given these potential explanations, we conducted three types of analysis to help diagnose the issue. Each analysis was designed to address a different set of potential explanations.

First, we retrieved data that we had collected and/or assembled (during the first phase of this research project) at 36 ODOT traffic signals for 48 hours each in 2022. These data included hourly counts (by phase) of the number of pedestrians observed (ground-truth measurements from video recordings), as well as the number of pedestrian services, pedestrian actuations, and unique pedestrian actuations (processed by USU/PSU from raw traffic signal controller log data). For a limited number of locations, we compared the daily totals and time-of-day profiles of our data to the data/estimates in the Portal data (for similar weekdays/months in the before and after periods). This allowed us to determine if data in the push-button database were roughly similar to those seen in 2022. The limitation here is that the 2022 data were for two particular days, so differences might not indicate an anomaly.

Second, we inspected how data were processed within Portal to check for any unexpected results. For each of a select number of locations (among those identified in Part I), we picked a random day in the before period and a random day in the after period. We extracted data from the *ped_data_raw* table, and tried to replicate the steps done to generate the data in the *ped_data_estimated* table. This included doing temporal aggregations (15-minutes to hours, hours to days) and applying the pedestrian volume estimation equations. This allowed us to determine if the data processing within Portal was happening as expected. The limitation here is that any issues with the data processing before the data reaches Portal cannot be determined.

Third, we inspected the extent of data completeness and treatment of missing data within the Portal push-button database and dashboard. For each of the intersections selected in Part I, we calculated and plotted the number of observations and the number of zero and non-zero (pedestrian activity) observations on each day in the before and after periods. This allowed us to see the extent of data completeness and how it might explain the anomalous volume trends we were seeing, as well as determining if the treatment of missing data within Portal was happening as expected. The limitation here is that we cannot determine how missing data are identified or zero data are treated, before reaching Portal.

Results 1: Comparisons to 2022 data

Of the 36 ODOT signals studied in 2022, only 18 were among the 585 intersections classified in Part I. We investigated four of these locations (each in a different anomaly category), as illustrated and described below.

E 1st St (OR-99W NB) and College St, Newberg (03051)

This signal was in the "mean lower, CV similar" category. For this signal, we assembled data for Thursdays, Fridays, and Saturdays in June. Figure 5 shows that estimated volumes and unique actuations in the after period roughly match data observed in 2022, whereas data in the before period are much lower and a poor match (phase 4) or are missing entirely (phase 8).

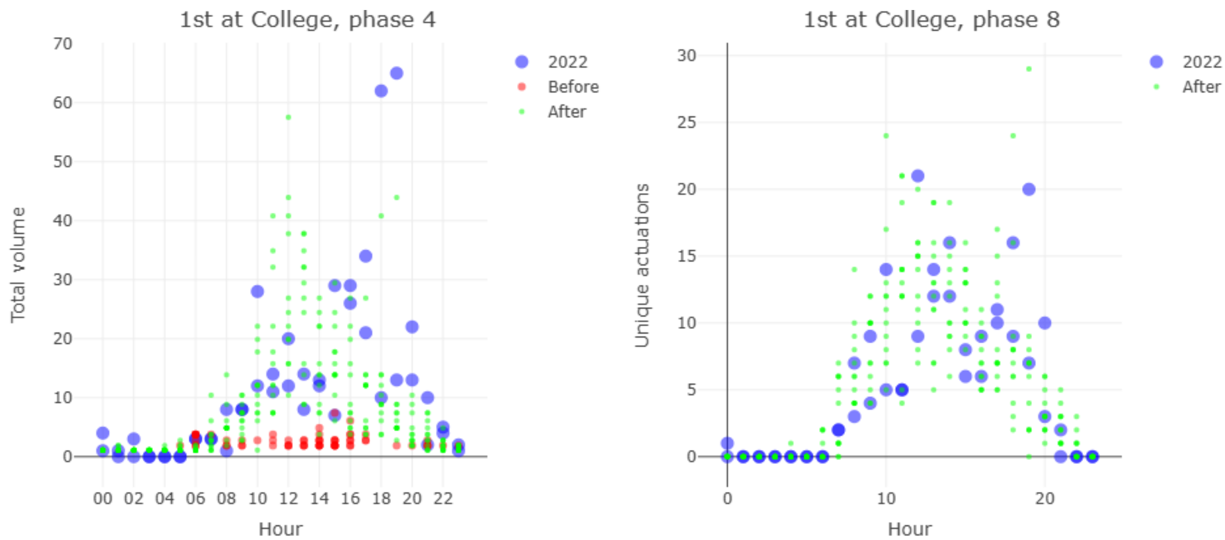


Figure 5: Comparisons of 2022 data to Portal data (before vs. after) at signal 03051

SE Powell Blvd (US-26) and SE 92nd Ave, Portland (2B100)

This signal was in the "mean lower, CV higher" category. For this signal, we assembled data for Tuesdays, Wednesdays, and Thursdays in June. Figure 6 shows that estimated volumes and unique actuations in the after period roughly match data observed in 2022, whereas data in the before period are much lower and a poor match.

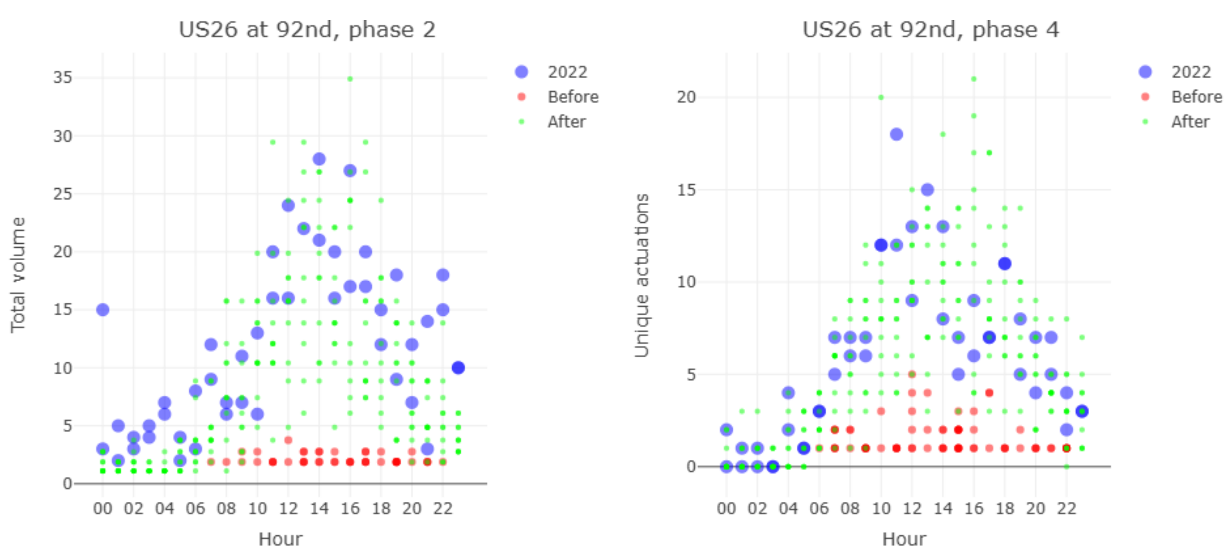


Figure 6: Comparisons of 2022 data to Portal data (before vs. after) at signal 2B100

SW Baseline St (OR-8 WB) and S 1st Ave, Hillsboro (2B404)

This signal was in the “mean lower, CV lower” category. For this signal, we assembled data for Tuesdays, Wednesdays, and Thursdays in June. Figure 7 shows that estimated volumes and unique actuations in the after period roughly match data observed in 2022, whereas data in the before period are lower and a poor match.

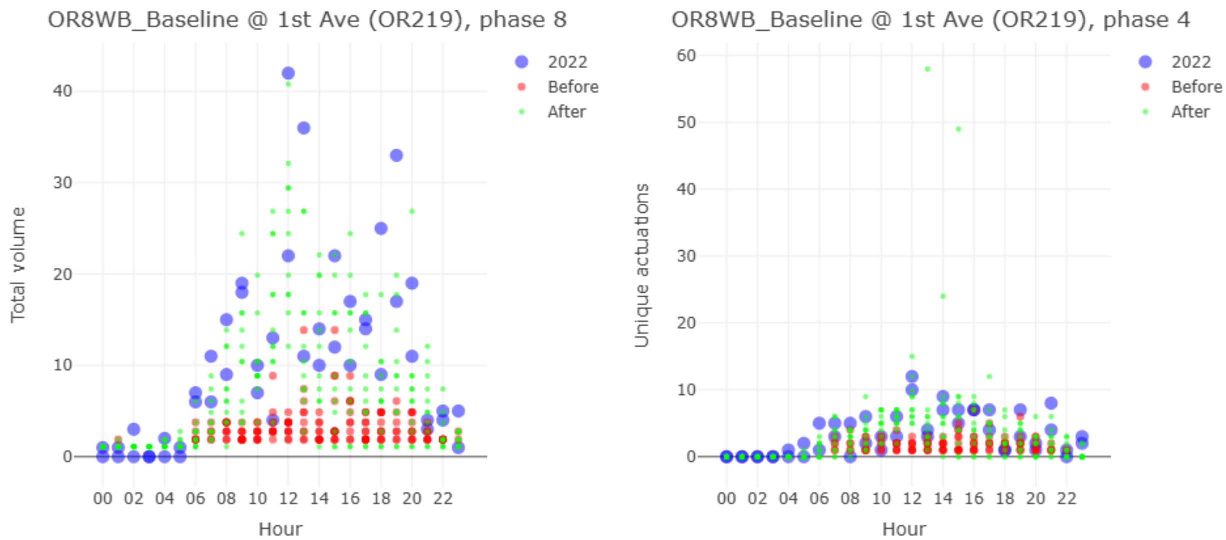


Figure 7: Comparisons of 2022 data to Portal data (before vs. after) at signal 2B404

Cornelius Pass Rd and US-26 EB Ramps, Hillsboro (2B471)

This signal was in the “mean higher, CV similar” category. For this signal, we collected data for Fridays, Saturdays, and Sundays in June. Figure 8 shows that estimated volumes and unique actuations in the after period roughly match data observed in 2022, whereas data in the before period are much higher and a poor match.

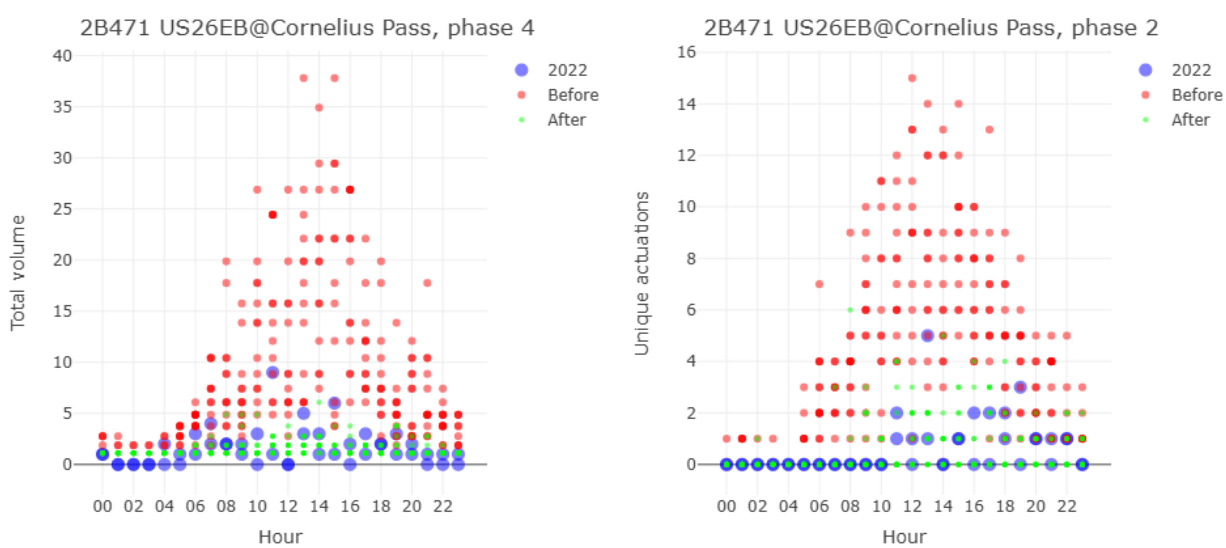


Figure 8: Comparisons of 2022 data to Portal data (before vs. after) at signal 2B471

Conclusions

At all four intersections, we found that estimated pedestrian volumes and unique pedestrian actuations during the after period (November 2024 thru October 2025) were much closer to 2022 observed pedestrian volumes and unique pedestrian actuations (by hour and phase) than those during the before period (June 2023 thru May 2024). This suggests that the **before data are more likely to be "wrong" or at least different**, for some reason. This analysis did not determine why before data are so much more different than after data, but subsequent analyses may be able to investigate this.

More specifically: The correspondence between observed 2022 and estimated (after period) pedestrian volumes suggest that (at least at these four locations), pedestrian behaviors did not change significantly and thus may not be the cause of the potential anomalies. Also, the correspondence between 2022 and after period unique actuations suggests that pre-Portal data processing (at least in the after period) could be accurate and thus also may not be the cause of the potential anomalies. Instead, there may be an issue with how data were processed (at some stage) during the before period.

Results 2: Data processing in Portal

We investigated seven intersections listed at the end of Part I (2B404, 2B349, 2B100, 04092, 13018, 01069, 03093), one from each anomaly category. For each intersection, we randomly selected a single day in each of the before and after periods. For each day, we processed the raw data using the follow steps, in an attempt to replicate the expected data processing:

- Get the estimated data (*ped_data_estimated*), for all (available) 1-hour time intervals and pedestrian phases. Also get the estimated data, for each phase, totaled for the day.
- Get the raw data, for all (available) 15-minute time intervals and pedestrian phases.
- Aggregate the raw data (*ped_data_raw*), by hour, for each phase. Sum the number of unique pedestrian actuations and the number of 15-minute observations.
- Apply the pedestrian volume estimation equations to the hourly aggregated raw data. (We used the equation for total pedestrian volume.)
- Compare the results at an hourly level (by phase) to the estimated hourly data: number of observations, unique actuations, total estimated volume.
- Aggregate the processed hourly raw data, by day, still for each phase.
- Compare the results at a daily level (by phase) to the estimated daily data: number of observations, unique actuations, total estimated volume.

For all seven intersections (and each phase at each signal), this calculation process was able to process the raw data and **perfectly replicate the results contained in the estimated data table**. Specifically, the hourly and daily totals of the "sample_count", "unique_actuations", and "total_estimated_volume" columns in *ped_data_estimated* were verified, at each intersection, for a random day in both the before and after periods.

Conclusions

Based on these findings, we conclude that the data processing within Portal—from the raw data (obtained from ODOT) to the estimated pedestrian volumes—is likely happening as expected. In other words, the data processing seems to be aggregating 15-minute raw data into hourly data, applying the pedestrian estimation equations, and (as necessary) aggregating the hourly estimates into daily estimates as we would expect. Data processing within Portal does not seem to be a primary cause of the potential anomalies.

Nevertheless, we should note that this conclusion is conditional upon the assumptions that (a) the raw data table is ingesting data from ODOT in a consistent manner across locations and over time, and that (b) the aggregation and estimation process is valid for the way the raw data are structured. It is possible that the way that data were being processed before reaching Portal changed, or that there are issues in the linkage of data to specific signals, or in the treatment of missing or incomplete data. We will investigate these possibilities in the following analysis.

Results 3: Data completeness

Based on the results of the previous two analyses, we narrowed in on the before period and the construction of the raw data table. We wanted to look for notable differences in the data contained in the *ped_data_raw* table during the before and after periods. Therefore, we investigated all 19 intersections identified in Part I for completeness and the treatment of missing data. Specifically, for each day in the before and after periods, we calculated the number of records—as well as the number of zero and non-zero counts of unique pedestrian actuations—for each phase at each intersection. We then plotted the results and visually inspected them for notable differences. The results are presented in Table 3, with several examples plotted and discussed below.

Table 3: Results of analysis of differences in data completeness

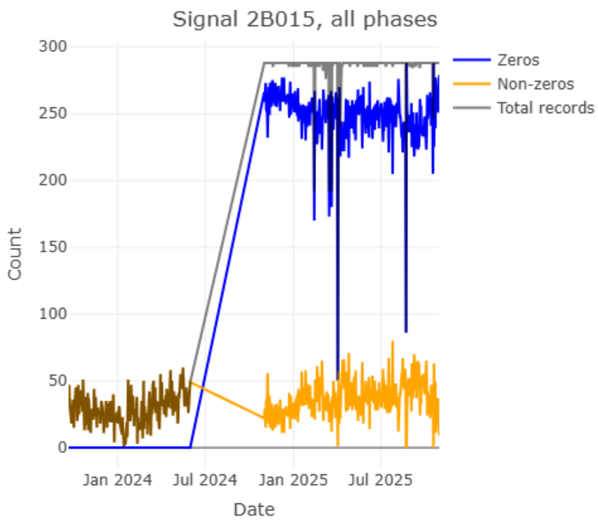
Anomaly category	***Intersection***	***Notable differences***
Mean lower, CV lower	Yew Ave / Airport Wy and US-97 NB Ramps, Redmond (10097)	Before period has phase 4, after period does not.
	SW Baseline St (OR-8 WB) and S 1st Ave, Hillsboro (2B404)	Before period has phase 2, after period does not. For all phases, after period has more non-zero records.
Mean lower, CV similar	Delta Hwy (OR-132) and Green Acres Rd, Eugene (05166)	Before period has phases 2 and 4, after period does not. For all phases, after period has more non-zero records.
	OR-99E and 10th St, Oregon City (2B015)	No notable observations.
	OR-224 and SE Monroe St, Milwaukee (2B050)	For phase 8, after period has only zeros.
	Pacific Hwy W (OR-99W) and St Main St and SW Johnson St, Tigard (2B349)	After period has phase 8, before period does not. For phase 2, after period has more non-zero records. For phase 4, before period has more non-zero records.

	State St (OR-43) and Foothills Rd, Lake Oswego (2B485)	No notable observations.
Mean lower, CV higher	Altamont Dr and Laverne Ave, Klamath Falls (11039)	Apparent structural difference: higher volumes in June-July of the before period.
	Adams Ave (US-30) and Fir St, La Grande (13004)	After period has phases 2 and 8, before period does not.
	SE Powell Blvd (US-26) and SE 92nd Ave, Portland (2B100)	For all phases, after period has more non-zero records.
Mean similar, CV lower	S Hwy 101 (US-101) and SE/SW 51st, Lincoln City (04092)	After period has phases 4 and 6, before period does not. For phase 2, before period has mostly zeros. For phase 8, before period has more non-zero records.
	Court Ave (US-30 WB) and Main St, Pendleton (12008)	Before period has phase 2, after period does not. For phase 8, after period has only zeros.
Mean similar, CV higher	E 2nd St (US-30 WB) and Washington St, The Dalles (09002)	No notable observations.
	Main St (US-30, OR-7) and Auburn Ave, Baker City (13018)	For phase 8, before period has more non-zero records.
Mean higher, CV similar	US-30 and Columbia Blvd, St Helens (01069)	For phase 6, before period has more non-zero records.
	US-199 and OR-99 SB, Grants Pass (08021)	Before period has phase 8, after period does not. For all phases, before period as more non-zero records.
	SW 4th Ave and SW 9th St and S Park Blvd, Ontario (14005)	Before period has phase 9, after period does not. For phase 6, before period as more non-zero records.
Mean higher, CV higher	N College St (OR-219) and E Mountainview Dr, Newberg (03093)	Before period has phases 12, 14, 16, and 18, after period does not. For most phases, before period has more non-zero records.
	Pacific Blvd S (OR-99E) and Allen Ln, Albany (04070)	Before period has phase 6, after period does not. For phases 2, 4, and 8, before period has more non-zero records.

OR-99E and 10th St, Oregon City (2B015)

State St (OR-43) and Foothills Rd, Lake Oswego (2B485)

Figure 9 shows the plots of completeness and zeros for these two intersections (2B015 and 2B485). As mentioned in Table 3, overall, there were no notable differences in the number of non-zero records per day between the before and after periods. However, notice how the total records equals the non-zero records in the before period, whereas the total number of records in the after period is relatively stable. In other words, **there are no zero records in the raw data table during the before period**. Notably, this observation was found for all signals investigated. However, during the after period, there seems to be complete records (including those with zero pedestrian actuations), as shown by the nearly-flat total records lines in the figures. There should be 96 complete records for each phase (24 hours × four 15-minute periods per hour), and $4 \times 24 \times N_{phases}$ number of complete records for each signal.



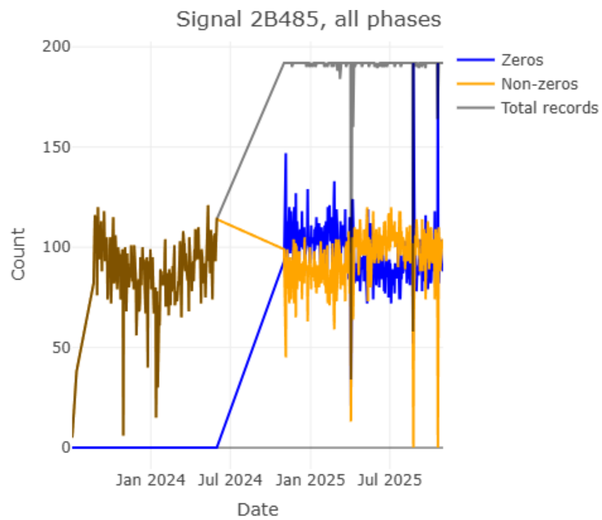


Figure 9: (a) Data completeness and zeros at signal 2B015. (b) Data completeness and zeros at signal 2B485.

Altamont Dr and Laverne Ave, Klamath Falls (11039)

SE Powell Blvd (US-26) and SE 92nd Ave, Portland (2B100)

Figure 10 shows the plots of completeness and zeros for these two intersections (11039 and 2B100). The individual phase plots showed similar trends. At the Klamath Falls intersection, the trends are relatively similar before and after, but the first two months (June and July) are much higher. Perhaps there was a significant change at this intersection around July or August 2023. At the Powell Blvd intersection, the after period has many more time periods with pedestrian activity than the before period. This large shift seems unlikely to be explained by behavioral or operational differences. It could be explained by **significant amounts of missing data before**, or an error in the matching or IDs of intersections; i.e., **the before period intersection is different than the after period intersection**.

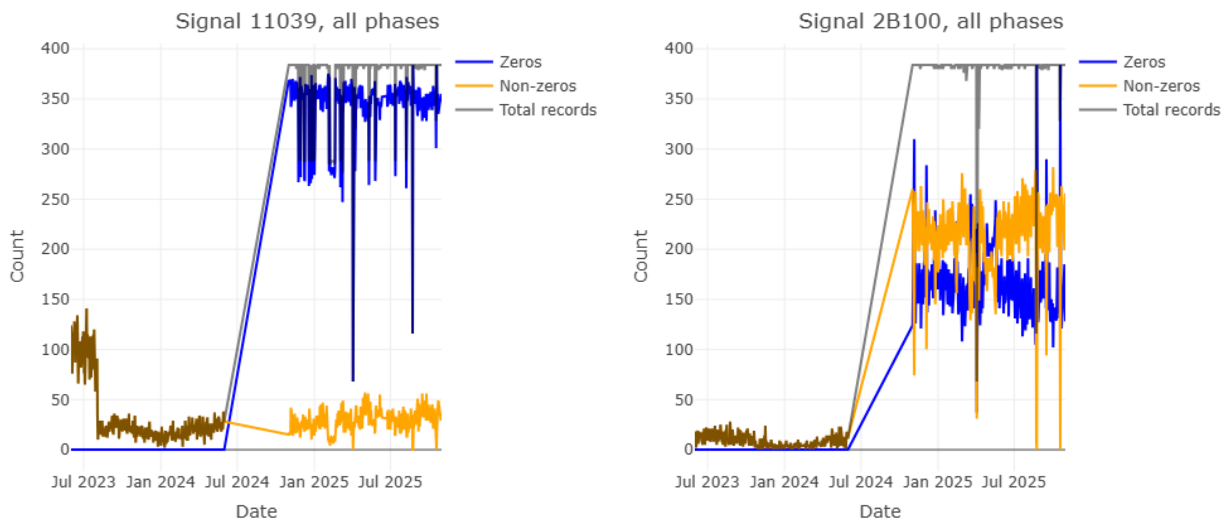


Figure 10: (a) Data completeness and zeros at signal 11039. (b) Data completeness and zeros at signal 2B100.

Delta Hwy (OR-132) and Green Acres Rd, Eugene (05166)

Figure 11 shows plots of completeness and zeros for phases 2 and 6 at this intersection (05166). The raw database has data for phases 2/4/6/8 in the before period, but only for phases 6/8 in the after period. Also, notice the large change for phase 6. Looking at this particular intersection, there are only two crosswalks, which raises a suspicion that the before data are coming from a different intersection.

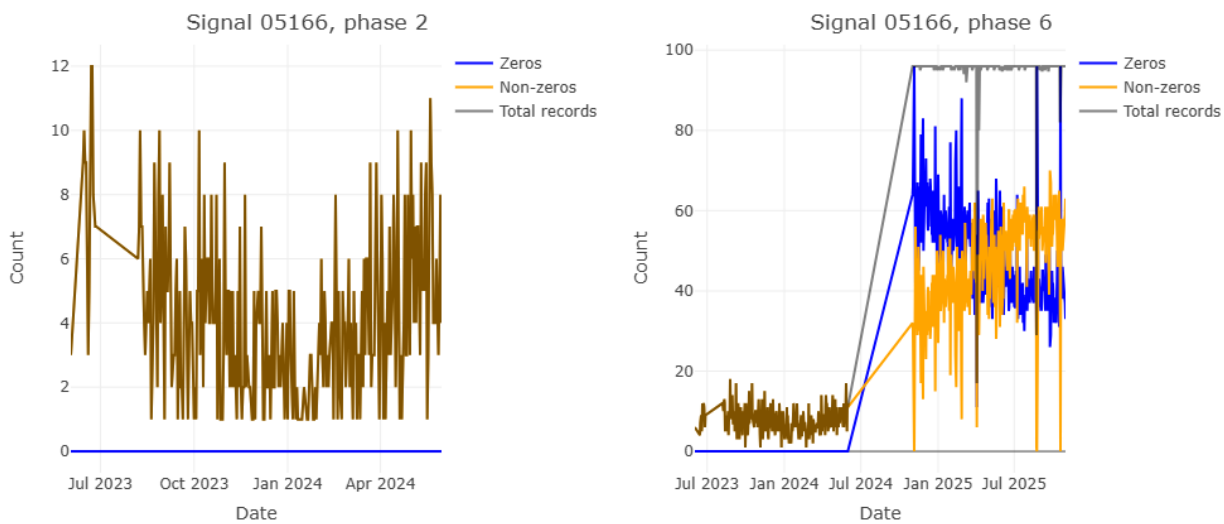


Figure 11: Data completeness and zeros at signal 05166 for phase 2 (left) and phase 6 (right).

OR-224 and SE Monroe St, Milwaukee (2B050)

Figure 12 shows plots of completeness and zeros for phases 4 and 8 at this intersection (2B050). Phase 4 shows a larger and more varied trend in the after period, while phase 8 has 100% zeros in the after period. We are unsure what might be causing this. One possibility is that the signal operation could have changed, putting both pedestrian crossings of the mainline (OR-244) on phase 4; although, this is pure speculation.

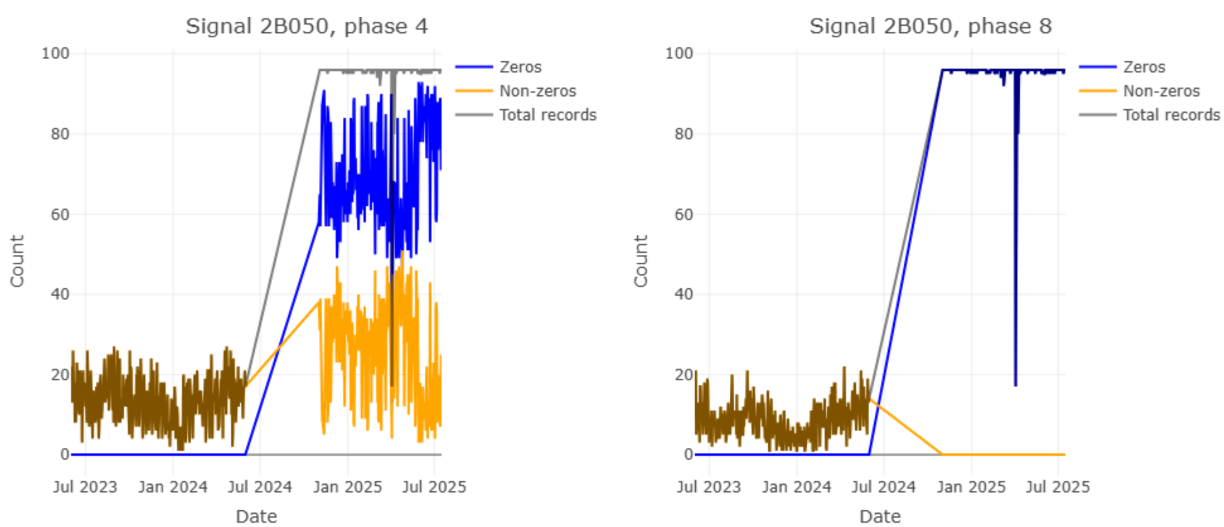


Figure 12: Data completeness and zeros at signal 2B050 for phase 4 (left) and phase 8 (right).

SW 4th Ave and SW 9th St and S Park Blvd, Ontario (14005)

Figure 13 shows plots of completeness and zeros for phases 6 and 9 at this intersection (14005). Phase 6 shows lower activity in the after period, and phase 9 shows no records in the after period. This intersection appears to be a standard four-leg intersection, so the presence of a ped phase 9 makes us suspicious that the before data are coming from a different intersection.

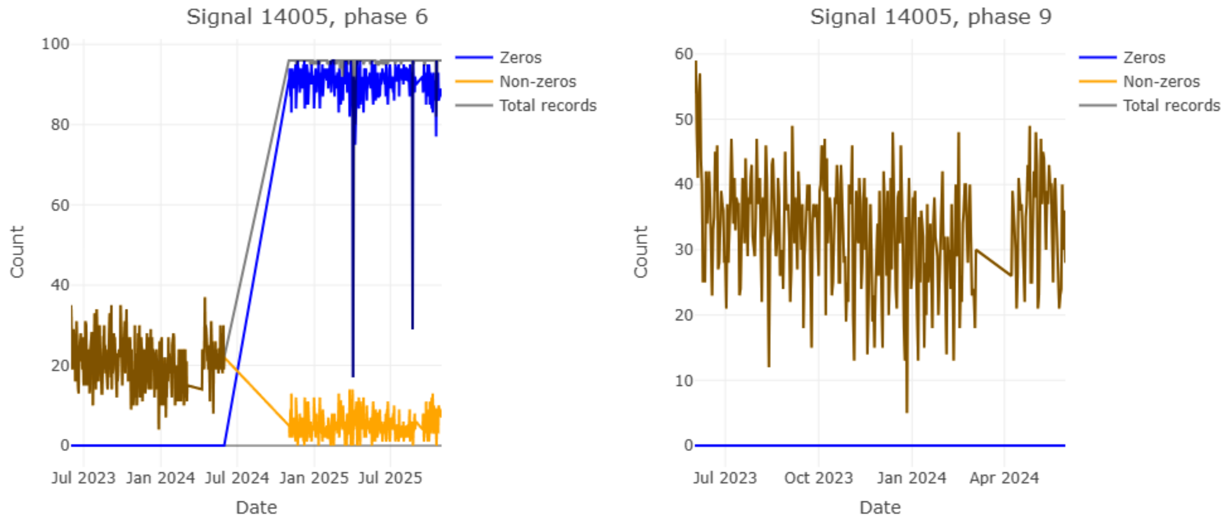


Figure 13: Data completeness and zeros at signal 14005 for phase 6 (left) and phase 9 (right).

Pacific Blvd S (OR-99E) and Allen Ln, Albany (04070)

Figure 14 shows plots of completeness and zeros for all phases and phase 8 at this intersection (04070). Plots for phases 2, 4, and 8 look similar to the all-phase plot. Phase 6 has no data in the after period. Given that this intersection has only three legs, and is located in an area where we would expect very low pedestrian activity, it seems likely that the data in the before period are coming from a completely different intersection.

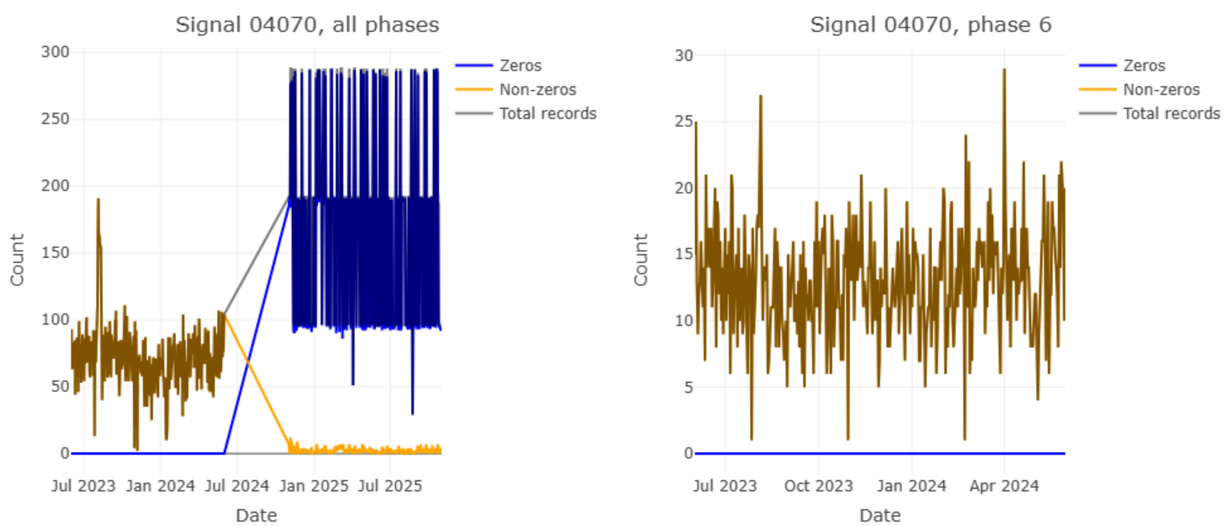


Figure 14: Data completeness and zeros at signal 04070 for all phases (left) and phase 8 (right).

Conclusions

This analysis yielded two important conclusions. *First*, there appears to have been **a major change in how the raw data were constructed** (prior to their ingestion into Portal). Before June 2024, records (15-minute periods, by intersection and phase) only entered the raw data table if there was some pedestrian signal activity; i.e., at least one unique pedestrian actuation. After October 2024, there was a record in the raw data table for (almost) every combination of intersection, phase, and 15-minute time interval, even if there was no pedestrian activity (i.e., a zero record). As a result, the before period has all of its zero records missing.

This finding implies that the data aggregation and pedestrian volume estimation process as currently working in Portal is not accounting for these missing zero (no pedestrian actuation) records during the before period. The reason is that the pedestrian volume estimation equations have a non-zero intercept term. Even when there are no unique pedestrian actuations (zero), the equation predicts around 1.1 pedestrians per hour. Therefore, it is likely that the before period estimated volumes are too low. Thus, this could explain why many of the intersections have lower estimated volumes during the before period (see Table 1). We will discuss potential solutions to this issue in the recommendations section below.

Second, at several if not many locations, there appears to have been structural changes in the relative number of 15-minute time intervals with non-zero pedestrian actuations, between the before and the after periods. In some locations, there were even pedestrian phases that were present in the data in one period but not the other. Together, this evidence could be explained by the data coming from different intersections in the two periods. In other words, **the device IDs in the databases may not be linked to the same (correct) intersections**. In the few instances we highlighted above, we found more reasons to question the before data and trust the after data; although, this may not be the case at all intersections. It is difficult to confirm this hypothesis or even evaluate the extent to which this may be the case, using just the data available in Portal. Trends can look similar but be from different intersections. But, if intersections were switched, this could explain many of the differences in trends that we identified in Part I.

Overall conclusions and recommendations

Overall, we investigated potential anomalies in estimated pedestrian volume data at signalized intersections in Oregon. After outlining the scope of the issue, we explored potential explanations and came to several important conclusions:

- In most cases, we did not find evidence that changes in pedestrian behavior or traffic signal operations can explain these anomalies. We also think that this is unlikely on a large scale.
- The data aggregation and pedestrian volume estimation processes within Portal are working as expected, although this is assuming a certain structure to the raw data table.
- There was a structural change in data within the *ped_data_raw* table. In the before period, only 15-minute periods with some pedestrian activity were included. In the after period, all 15-minute periods (even those with no pedestrian activity) were included. This change means that most the before period volume estimates are likely too low.

- There is evidence that, at least at some intersections, the data in the before period may be from different intersections than the data in the after period. If true, this means that the before period data at these locations cannot be compared to the after period data.

Based on those conclusions, we have the following recommendations:

1. **Review intersections and their device IDs through the data processing pipeline.** Verify that the current information in the *intersections* table is correct. Check that the device IDs for data prior to mid-2024 link to the same intersections now (since late 2024). (We found more evidence that the after-period intersections are correct, and that some before-period intersections are incorrect.) This likely requires investigation by ODOT staff about how raw signal data are processed, with coordination by PSU staff about how data are ingested into Portal push-button database tables.
2. **Discuss what to do with data before November 2024.** This depends on the conclusions of the first recommendation above.
    a. If intersections were switched (but are correct now) and fixes cannot be made, we think that any data prior to late 2024 should be discarded as unreliable.
    b. If intersections were switched and fixes can be made, those adjustments should probably be made. However, in this case, we suspect that the raw *ped_data_raw* table will still have structural differences (before late 2024) that may prove problematic. Specifically, during the before period, we don't know if the lack of a record in the table is because there was no pedestrian activity (zero record) or because the data were truly missing (unknown pedestrian activity) for either the phase or the intersection as a whole. If a true zero, the data aggregation process should account for this and estimate a volume for this period. If a true missing record, the data process should account for this the same way it accounts for missing data (see next recommendation). Perhaps the *has_data* table could help with this determination; although we did not review that table. Information about its construction would be helpful to inform this recommendation.
3. **Clarify the treatment of missing data in the push-button database and dashboard.** Even during the after period, some records are missing for various reasons. The dashboard should make it clear to users how missing data are treated. We assume that Portal has standard missing data procedures and information for other data types (show percent complete, allow imputation, etc.), which may be applied to pedestrian push-button data in similar ways. Time periods with (or aggregations over) significant quantities of missing data may need to be flagged or omitted.

Please contact me with any questions, including suggestions for improvements to this memo. Thank you for the opportunity to support the implementation of this research and the dissemination of pedestrian volume data for multifaceted important uses.

Patrick A. Singleton, Ph.D.
Associate Professor
Department of Civil & Environmental Engineering
Utah State University, Logan, UT
patrick.singleton@usu.edu | 435.797.7109
https://engineering.usu.edu/cee/people/faculty/singleton-patrick