

Aim

The aim of this project is to build an automated SPC (Statistical Process Control) engine and dashboard that monitors vaccination progress across regions and population groups. By comparing group wise real-time data with historical trends, the system detects abnormal behaviour, identifies early warning signals, and highlights regions that require intervention.

How the Algorithm works

A. Build Baseline (from Autumn 2024 + Spring 2025)

1. Load historical CSVs
2. Convert dates, clean entries
3. Compute weekly vaccination rate for each region/group
4. For each week number × group, calculate:
 - a. expected weekly rate (mean across regions)
 - b. sigma = natural variation (standard deviation)
5. Save this baseline as the model

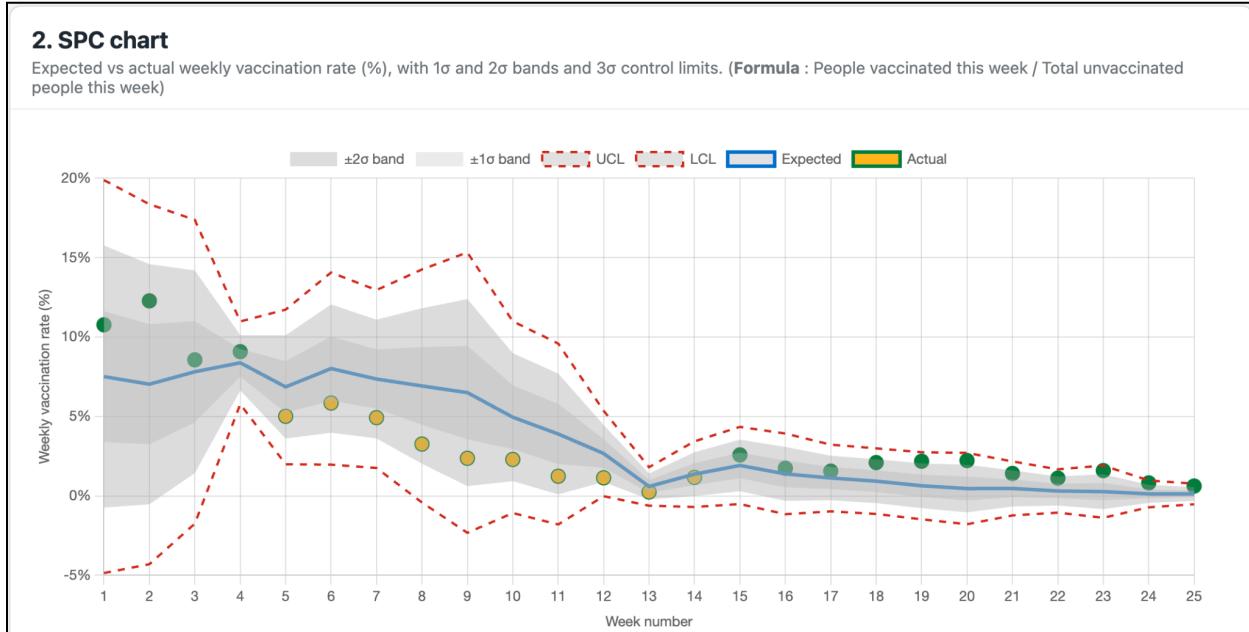
B. When a NEW campaign CSV is uploaded

1. Preprocess the data
 - Sort
 - Remove duplicates
 - Parse dates
2. Compute weekly vaccination rate
 - new_vacc = change in vacc_alive from last week
 - remaining_unvaccinated = denom - previous week's vacc_alive
 - rate = new_vacc / remaining_unvaccinated
3. Attach baseline
 - For each row, attach expected_rate and sigma_rate from the baseline model
4. Compute SPC components
 - z-score
 - Upper control limit (UCL = expected + 3 σ)
 - Lower control limit (LCL = expected - 3 σ)
 - Status = Green / Yellow / Red (Calculated using the z-score)
5. Apply SPC run rules
 - 8 points below expected
 - 2 of last 3 below -2 σ
 - 6-week trend
6. Build outputs
 - Wales-wide aggregated SPC curve
 - Dashboard summary

- Deviation table (points outside normal range)
7. Send results to frontend
- JSON → rendered using Chart.js & HTML tables

How to Read the output

A. SPC Chart (Main Graph)



- Blue line = expected vaccination progress
- Grey band = natural variation (1σ and 2σ ranges)
- Red dashed lines = UCL/LCL (3σ)
- Green/yellow/red dots = actual weekly vaccination rate
 - Green: Normal
 - Yellow: Below expected
 - Red: Outside 3σ → needs investigation
- Use this to spot early drops in performance

B. Dashboard Summary (Affected by Region and Group filters)

3. Dashboard summary

Counts of weeks in green, amber and red, ordered by where problems appear most often.

Region	Group	Total weeks	Green	Yellow	Red	Run rule alerts
All Wales	All identified as potentially eligible	25	11	14	0	4
All Wales	Residents in a care home for older adults	25	10	11	0	2
All Wales	Adults aged 65 years and over	25	15	10	0	4
All Wales	Persons aged 6 months to 64 years in a clinical risk group	25	15	10	0	0

Shows for each region & group:

- total weeks
- number of green weeks
- number of yellow weeks
- number of red weeks
- number of run rule alerts* weeks

Sorts by:

1. Highest red weeks
2. Highest yellow weeks
3. Highest green weeks

This helps identify which regions need immediate operational attention.

C. Deviation Table

4. Major deviations from expected

All weeks where regions/groups are much lower or higher than we'd expect, based on the Wales-wide baseline.

Region	Group	Week end	Weekly rate	Expected	Sigma	Classification
Aneurin Bevan UHB	Residents in a care home for older adults	12 Jan 2025	18.0%	1.8%	4.7%	Alert - extremely higher than expected
Aneurin Bevan UHB	Residents in a care home for older adults	19 Jan 2025	2.2%	0.3%	0.6%	Alert - extremely higher than expected
CTM UHB	Residents in a care home for older adults	01 Dec 2024	12.7%	2.5%	3.2%	Alert - extremely higher than expected
Cardiff and Vale UHB	Residents in a care home for older adults	26 Jan 2025	20.0%	1.9%	5.0%	Alert - extremely higher than expected
Hywel Dda UHB	Persons aged 6 months to 64 years in a clinical risk group	01 Dec 2024	9.8%	1.7%	0.8%	Alert - extremely higher than expected

Lists **ONLY** abnormal points, classified as:

- Lower than expected (1σ - 2σ)
- Higher than expected (1σ - 2σ)
- Extreme → Below LCL / Above UCL

Use this table to quickly identify specific weeks where a major change happened.

**Run rule Alerts :*

- Rule 1 : 8 Consecutive Points Below Expected
- Rule 2 : 2 Out of 3 Weeks More Than 2σ Below Expected
- Rule 3 : 6 Consecutive Weeks Trending Up or Down

XXX

Code Explanation

Important functions used in app.py

- SPConfig
Stores global settings such as:
 - ❖ column names
 - ❖ σ thresholds
 - ❖ run rule window lengths
 - ❖ min/max weekly rate limitsThis keeps the engine flexible.
- fit_baseline_model(history_df, cfg)
Uses historical campaigns to compute:
 - ❖ expected_rate
 - ❖ sigma_ratefor each week_num \times group.
This acts as the "gold standard" that new data is compared against.