Filled posts monthly tracking: revision notes

2024/25

Contact us: analysis@skillsforcare.org.uk

Contents

Timeline

Version 1: 15th November 2024

- First version using our data engineering pipeline published for care homes.

Version 2: 13th December 2024

- Percentage changed added for non-residential locations in 2024.
- Improvements made to the rolling average model.

Version 3: 23rd June 2025

- Included the estimated number of filled posts for non-residential locations in the dashboard.
- Set a maximum interpolation period.
- Model improvements (changed to linear regression and feature improvements).

Version 4: 15th December 2025

- Amended rate of change calculation to match new interpolation methods in version 3.
- Delayed forwards extrapolation to match the median number of months locations stay the same.

Version 5: 13th March 2026

- Replaced the static care home and non-residential with/without dormancy with models which re-train with the latest data.
- Amended the delaying of forwards extrapolation introduced in version 4 to take workplace size into account.
- Added filtering of non-residential outliers in the rate of change calculations.
- Changed imputation to apply nominal changes instead of applying a ratio.

Introduction

To predict the monthly changes in filled posts within adult social care, we first estimate the number of filled posts at each independent CQC-regulated location. For a detailed methodology, see the Methodology page of our Workforce Intelligence website.

Version 2

Improvements to the rolling average model

Independent sector users of ASC-WDS submit data at varying frequencies, from monthly to annually. This results in gaps where data is not submitted for certain months, see Table 1. To address these gaps, we generate a trendline to extrapolate forwards or backwards and interpolate between known values. The way we generate this trendline has changed in version two of these estimates.

Table 1. Example data for illustration

Source: Example data

Location	Month 1	Month 2	Month 3	Month 4
Location 1		25.0		
Location 2	50.0		51.0	
Location 3		76.0		77.0
Location 4	100.0		102.0	
Location 5	125.0		128.0	
Location 6		151.0		155.0
Location 7	175.0		179.0	180.0
Location 8		202.0		206.0

Version 1: Six-month rolling average trendline

Initially, we calculated a monthly average of all submitted data and then applied a six-month rolling average. However, infrequent submissions from atypical locations (for example very small or large numbers of staff) could skew the monthly average, leading to trends that were potentially not representative of the whole sector.

Table 2. Monthly averages and changes based on illustrated data in Table 1

Source: Example data

	Month 1	Month 2	Month 3	Month 4
Monthly average	112.5	113.5	115.0	154.5
Change since previous month		0.9%	1.3%	34.3%

As Table 2 shows, the change in month four is very large and is caused by larger than average locations submitting in that period. But looking at their previous submissions, they were only growing at a modest rate. So, applying a large rate of growth to other locations based on this data would not be accurate.

The six-month rolling average would smooth out individual monthly spikes to some extent,

but they would still have an impact.

Version 2: Rate of change trendline

Our solution was to focus more on how locations are changing from one month to the next, as opposed to top level averages. The first step was to remove locations who had submitted only once (in red in Table 3) and to fill gaps between submissions using a straight-line imputation approach (in purple in Table 3). This updates the original data in Table 1 to the following dataset.

Table 3. Original data imputed with straight-line interpolation and single submissions removed

Source: Example data

Location	Month 1	Month 2	Month 3	Month 4
Location 1		25.0		
Location 2	50.0	50.5	51.0	
Location 3		76.0	76.5	77.0
Location 4	100.0	101.0	102.0	
Location 5	125.0	126.5	128.0	
Location 6		151.0	153.0	155.0
Location 7	175.0	177.0	179.0	180.0
Location 8		202.0	204.0	206.0

The next step is to calculate an individual monthly rate of change. A location only qualifies as being included in the monthly rate of change if they have a known value in that specified month and the previous month. We then sum the values of all the qualifying locations for each month and the previous month to get the overall rate of change of all those locations combined, see Table 4.

Table 4. Sum of locations who qualify for rate of change method using illustrated data from Table 3

Source: Example data

	Month 1	Month 2	Month 3	Month 4
Sum of values (specified month)	450.0	455.0	893.5	618.0
Sum of values (previous month)	-	450.0	884.0	612.5
Change since previous month	-	101.1%	101.1%	100.9%

As before, we take the six-month average change into account to smooth out the trendline.

We found this trendline reflects monthly changes more accurately than the overall average because it is less affected by atypical locations joining, leaving, or not submitting data from one month to the next.

Impact on published figures between version 1 and 2

Table 5 below shows the impact on the estimates between the two sources. Note that these are presented as unrounded numbers here to assess the scale of change. When published they are rounded to reflect the fact they are estimates and not counts.

Table 5. Comparison of estimates for care homes by version
Source: Skills for Care estimates

	Version 1	Version 2	Difference
Mar-24	589,178	575,630	13,548
Apr-24	591,051	576,255	14,796
May-24	591,461	576,072	15,389
Jun-24	587,837	577,172	10,665
Jul-24	589,590	579,622	9,968
Aug-24	587,566	579,813	7,754
Sep-24	581,690	582,715	-1,025
Oct-24	586,254	581,223	5,031
Nov-24		584,043	

Version 3

Set a maximum interpolation period

When data has been submitted in ASC-WDS more than once for a location, we estimate the values in between by assuming a straight line between the known points. These interpolated values are then used to calculate a trendline of how the workforce changes over time.

In version 2, all gaps between submitted values were filled, no matter how far apart the submissions were. However, this often reduced real changes in workforce numbers. For example, during the Covid-19 period when workforce levels shifted rapidly, straight-line estimates across long gaps (e.g. 12 months) flattened out meaningful trends.

In version 3, we introduced a 6-month maximum gap before calculating the rate of change trendline. This means we now only include submitted values that are no more than 6 months apart when calculating month-to-month changes. Shortening the interpolation window in this way makes the trendline more responsive and better able to reflect genuine shifts in the data.

Model improvements

Linear regression models (care home and non-residential models)

We have changed the care home and non-residential models from Gradient Boosted Trees (GBT) to linear regression models. The GBT models were overfitting and unstable at location level whereas the linear regression models offer better explainability and more stable trends.

Changed dormancy from a binary feature to scale (non-residential models only)

Locations in the CQC register are flagged as dormant if the location is not currently providing regulated services. This can happen when a service is undergoing renovation, newly opened and trying to win new business or currently only delivering non-regulated activities.

In March 2024 there were 1,854 dormant non-residential locations. By March 2025, 873 of these locations were no longer dormant.

In version 2, the model included dormancy as a binary Yes/No feature and each location coming out of dormancy accounted for a sudden step-change of around 30 additional filled posts on average.

Through discussion with the CQC and a review of other data sources, it was concluded that a change in dormancy status did not have such an immediate and large impact on workforce

size. As such, the current models were over-estimating growth in the non-residential sector between 2024 and 2025 when these locations were coming out of dormancy.

For version 3 the model uses a 'time since dormancy' (measured in months) instead of the binary yes/no dormancy flag. This resulted in a smoother transition, in terms of workforce size, when locations come out of dormancy. This is a more accurate reflection of what happens in practice based on the available evidence.

Added related location as a feature (non-residential models only)

The non-residential models use the number of months since the CQC location was registered as a feature. There are various scenarios where a location will be re-issued a new location ID and registration date, such as a location move or the location being taken over by a different provider. There is another field in the dataset which highlights if the current location ID was previously registered under a different ID number. Adding this 'related location' term as a feature in the model helped to better distinguish genuinely new services (who tend to have very few staff) from previously registered ones (who tend to be more established with more staff).

Impact on published figures between version 2 and 3

Table 6. Comparison of percentage change since March 24 for non-residential locations by version

Source: Skills for Care estimates

	Jun-24	Sep-24	Dec-24	Mar-25
Version 2 (percentage change)	2.4%	5.0%	6.8%	10.7%
Version 3 (percentage change)	1.5%	3.4%	3.3%	5.1%
Difference (percentage points)	-0.9	-1.6	-3.5	-5.6

Table 7. Comparison of percentage change since March 24 for care homes by version

Source: Skills for Care estimates

	Jun-24	Sep-24	Dec-24	Mar-25
Version 2 (percentage change)	0.4%	0.7%	1.5%	1.8%
Version 3 (percentage change)	0.7%	1.9%	2.5%	3.2%
Difference (percentage points)	0.3	1.2	1.0	1.4

Version 4

Amend rate of change calculation to match new interpolation

We calculate the monthly rate of change in filled posts to feed most of our models, including extrapolation, and regression models.

The algorithm used to calculate the rate of change in filled posts used to look at the last known value in our interpolated data and compared this to the current value in our interpolated data to determine the rate of change at a particular point in time.

As we used to interpolate all gaps between known submissions, this effectively looked at the previous month's data. As of version 3, we stopped interpolating between values that were more than six months apart to make the trendline more responsive. Because this leaves gaps in the interpolated data, using the last known value was no longer always getting the previous month's data. This caused the estimates for some locations to increase in short period of time rather than smoothly over the period between their updates. This has been corrected so the rate of change is now always explicitly calculated from the previous month's data.

Impact on published figures due to rate of change

Table 8. Comparison of percentage change since March 25 for non-residential locations by version

Source: Skills for Care estimates

	Version 3 (percentage change)	With rate of change amendment (percentage change)	Difference (percentage points)
Mar-25	0.0%	0.0%	0.0%
Apr-25	0.3%	0.1%	-0.2%
May-25	0.7%	0.2%	-0.5%
Jun-25	1.5%	0.6%	-0.9%
Jul-25	1.5%	0.4%	-1.1%
Aug-25	1.9%	0.6%	-1.3%
Sep-25	2.9%	1.2%	-1.7%
Oct-25	3.5%	1.4%	-2.1%

Table 9. Comparison of percentage change since March 25 for care homes by version
Source: Skills for Care estimates

	Version 3 (percentage change)	With rate of change amendment (percentage change)	Difference (percentage points)
Mar-25	0.0%	0.0%	0.0%
Apr-25	0.4%	0.3%	-0.1%
May-25	0.4%	0.3%	-0.1%
Jun-25	0.6%	0.5%	-0.1%
Jul-25	0.8%	0.6%	-0.2%
Aug-25	0.8%	0.6%	-0.2%
Sep-25	1.1%	0.8%	-0.3%
Oct-25	1.3%	0.9%	-0.4%

Delay forwards extrapolation

Extrapolation is used to estimate the number of filled posts at locations that have previously submitted data to ASC-WDS but have not provided a recent update.

In earlier versions of the estimates, extrapolated values were calculated using a rate of change derived from locations that had reported changes in filled posts between periods. This approach assumed that workforce size was changing continuously and did not account for locations that remain the same size for extended periods. As a result, the rate of change could be overstated.

To better understand how frequently workforce sizes change in practice, analysis was carried out using data from the Department of Health and Social Care's Capacity Tracker (CT). CT collects workforce information monthly from a large proportion of adult social care providers and indicated that the median length of time a location remains at the same workforce size is approximately three months.

To reflect this behaviour, version 4 introduced a delay before forward extrapolation is applied. The last submitted value is now carried forward for two additional months (meaning the same value is retained for up to three months) before the rate of change trendline is used to generate further estimates.

This change reduces the likelihood of introducing artificial increases or decreases where the workforce size has remained stable.

Impact on published figures due to delaying forwards extrapolation

Table 10. Comparison of percentage change since March 25 for non-residential locations by version

Source: Skills for Care estimates

	With rate of change amendment (percentage change)	With rate of change amendment and delaying forwards extrapolation (percentage change)	Difference (percentage points)
Mar-25	0.0%	0.0%	0.0%
Apr-25	0.1%	0.1%	0.0%
May-25	0.2%	0.4%	0.2%
Jun-25	0.6%	0.6%	0.0%
Jul-25	0.4%	0.7%	0.3%
Aug-25	0.6%	0.8%	0.2%
Sep-25	1.2%	0.9%	-0.3%
Oct-25	1.4%	1.0%	-0.4%

Table 11. Comparison of percentage change since March 25 for care homes by version
Source: Skills for Care estimates

Source: Skills for Care estimates

	With rate of change amendment (percentage change)	With rate of change amendment and delaying forwards extrapolation (percentage change)	Difference (percentage points)
Mar-25	0.0%	0.0%	0.0%
Apr-25	0.3%	0.3%	0.0%
May-25	0.3%	0.3%	0.0%
Jun-25	0.5%	0.5%	0.0%
Jul-25	0.6%	0.5%	-0.1%
Aug-25	0.6%	0.6%	0.0%
Sep-25	0.8%	0.8%	0.0%
Oct-25	0.9%	0.9%	0.0%

Version 5

Automated model retraining

In previous versions of the estimates, the regression models used to predict filled posts were trained periodically and then saved for use in the pipeline. The most recent training took place in June 2025, meaning subsequent pipeline runs continued to use the same model parameters despite additional workforce data becoming available.

From version 5 onwards, the models are retrained automatically each time the pipeline is run. This ensures the models learn from the most recent available data before generating predictions.

Automated retraining allows the estimation process to adapt to emerging workforce trends and reduces the risk that model predictions become less representative of current workforce patterns as new ASC-WDS data is collected.

Applying this change has resulted in a revision to recent estimates. Much of this revision reflects the previous models being trained on older data and therefore becoming less representative of more recent workforce patterns. Retraining the models using the latest available data improves the alignment between model predictions and the most recent observed submissions.

For non-residential locations, two separate models are used. A new feature indicating whether a location is recorded as dormant was introduced in the Care Quality Commission (CQC) data from March 2022. Incorporating this feature improves model accuracy, so a model including the dormancy variable is used for periods where this information is available.

As the historical ‘without dormancy’ model is only used to estimate earlier periods, its training window is fixed and it will not be retrained in future pipeline runs. The newer ‘with dormancy’ model will continue to be retrained using the latest available data.

Delay forwards extrapolation (size-based refinement)

The delay before forward extrapolation introduced in version 4 applied a single threshold across all locations. Further analysis showed that smaller locations are more likely to remain unchanged for longer periods, while larger locations tend to experience workforce changes more frequently.

In version 5, the delay before extrapolation is therefore adjusted according to workplace size. Smaller locations retain their last reported value for longer periods before extrapolation begins, while larger locations transition to extrapolated estimates sooner.

Filtering outliers in the rate of change calculations (non-residential locations)

The rate of change trendline introduced in version 2 uses month-to-month changes in filled posts to estimate workforce growth or decline. During further review of this data, a small number of observations were identified where the change between months appeared implausibly large.

These extreme values disproportionately influenced the calculated distribution of changes, particularly for non-residential locations, and could lead to exaggerated increases or decreases when imputing missing values.

Version 5 introduces an additional quality filtering step when calculating the rate of change for non-residential locations. Month-to-month changes are compared against the distribution of observed changes across the dataset, and values with unusually large absolute or percentage changes are excluded from contributing to the rate-of-change calculation.

These observations are not removed from the dataset and the original reported values remain unchanged within the pipeline. The filtering step only prevents atypically large changes from influencing the calculation of the typical changes applied to other locations.

Very small locations are excluded from this filtering step because small numerical changes can produce large percentage changes that remain plausible.

Excluding these extreme values improves the stability of the estimated rate of change and reduces the influence of unusual reporting behaviour at a small number of locations.

Applying nominal changes from model predictions during imputation

Previous versions of the estimates used the ratio between consecutive model predictions when imputing missing values between known data points. The ratio-based approach was originally intended to account for the expectation that larger locations may experience larger proportional workforce changes.

Further testing showed that applying the nominal change (the difference between consecutive model predictions) produced estimates that were consistently closer to observed values when compared with known data submissions. Version 5 therefore applies nominal changes from model predictions when adjusting values between known data points.

During intermediate calculations this approach can produce temporary negative values when the model predicts a decline. These values are retained during the calculation process so that decreases in workforce size are applied proportionally when adjusting from known values.

As in previous versions, a minimum value is applied at the final stage of the process to ensure that published estimates remain within plausible bounds.

Impact on published figures between version 4 and 5

Table 12. Comparison of percentage change since March 25 for care homes by version
Source: Skills for Care estimates

	Percentage change since March 2025		Percentage point difference
	Version 4	Version 5	
Apr-25	0.3%	0.3%	0.0%
May-25	0.3%	0.3%	0.0%
Jun-25	0.5%	0.5%	0.0%
Jul-25	0.6%	0.6%	0.0%
Aug-25	0.6%	0.6%	0.0%
Sep-25	0.8%	0.8%	0.0%
Oct-25	0.9%	0.9%	0.0%
Nov-25	1.2%	1.1%	-0.1%
Dec-25	1.3%	1.2%	-0.1%
Jan-26	1.3%	1.2%	-0.1%

Table 13. Comparison of percentage change since March 25 for non-residential locations by version
Source: Skills for Care estimates

	Percentage change since March 2025		Percentage point difference
	Version 4	Version 5	
Apr-25	0.1%	0.2%	0.1%
May-25	0.2%	0.6%	0.4%
Jun-25	0.5%	1.1%	0.6%
Jul-25	0.4%	1.3%	0.9%
Aug-25	0.5%	1.5%	1.0%
Sep-25	1.0%	2.1%	1.1%
Oct-25	1.0%	2.2%	1.2%
Nov-25	1.3%	2.4%	1.1%
Dec-25	1.4%	2.5%	1.1%
Jan-26	1.4%	2.7%	1.3%

Skills for Care
West Gate
6 Grace Street
Leeds
LS1 2RP

T: **0113 245 1716**
E: **info@skillsforcare.org.uk**

skillsforcare.org.uk

facebook.com/skillsforcare

linkedin.com/company/skills-for-care

© Skills for Care 2025