skillsforcare

workforce intelligence

Filled posts monthly tracking revision notes

2024/25

Contents

skillsforcare workforce intelligence

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)

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 here.

Version 2 changes

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 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

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

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

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

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

skillsforcare.org.uk

twitter.com/skillsforcare

facebook.com/skillsforcare

linkedin.com/company/skills-for-care