Revisions history

Contents

Technical detail

1. Our previous methods

Prior to automating our estimates of the number of filled posts in the CQC-regulated independent sector, we were only able to provide annual estimates of the size of the sector and monthly updates on the subset of the sector who were updating ASC-WDS during the year.

The monthly tracking was not able to account for new locations opening or locations closing during the year. Additionally, we were only able to report on percentage change since the previous annual estimates; we were unable to produce a whole sector estimate each month.

2. Our new methods

Our new automated method uses all available ASC-WDS data over time and estimates the number of filled posts for each location at the end of each month, using one of three methods:

A. **Using up to date ASC-WDS data**: If we have ASC-WDS data completed on the number of filled posts for a particular location in a particular month then that figure is used (after some data quality filtering).

B. **Using ASC-WDS data completed/updated during a different time period**: If we have ASC-WDS data for a particular location but it is not completed/updated in the month we are estimating, then we create an estimate based on data we have in the ASC-WDS for that location at different points in time (imputation).
   - We make estimates from the most recent known value at a location forwards in time and the earliest known value for that location backwards in time (extrapolation).
   - We make estimates for months between two known ASC-WDS values (interpolation).
   - Both extrapolated and interpolated estimates follow the trend seen in ASC-WDS data completed each month. A simplified example would be that if locations completing ASC-WDS in February 2024 were 1% larger, relative to their number of beds, than locations completing ASC-WDS in January 2024, then a location completing ASC-WDS data in January 2024 would have an extrapolated value for February 2024 that was 1% higher than the value they submitted in January 2024.

C. **Using a regression model:** For **non-residential locations only**, if we do not have ASC-WDS data for a particular location at all, but we do have a figure of people directly employed from the CQC Provider Information Return (PIR) dataset, then we use a linear regression model to create an estimate of filled posts for that location based on the CQC data.

D. **Using a machine learning model**: If we do not have ASC-WDS data for a particular location at all (or in the case of non-residential locations, no ASC-WDS or PIR data, see Accuracy), then we use a machine learning model to create an estimate of filled posts for that location based on data we have in the ASC-WDS for similar locations.
   The care home machine learning model uses the following features about a location to estimate the number of filled posts:
   - the number of beds
   - the number of services provided
   - the rolling rate of change of the filled posts to beds ratio at the time
   - the types of services provided
   - the geographical region
   - the ONS classification of rural-urban indicators.

   The non-residential machine learning models use the following features about a location to estimate the number of filled posts:
   - the number of activities provided
   - the number of services provided
   - the number of specialisms provided
   - the length of time the location has been registered
   - the rolling rate of change of the filled posts for non-residential locations at the time

- the types of services provided
- the geographical region
- the ONS classification of rural-urban indicators.
- whether the location was dormant at the time (this only exists for more recent data)

Combining the estimates for each location at each point in time allows us to provide an estimate for the number of filled posts each month for every

4. Accuracy

We run diagnostic checks on our regression model estimates and

Additionally, our estimates are never used at location level, they are always aggregated into geographical areas. So, an estimate falling outside of these metrics does not necessary cause an issue. The size and direction of the differences for each model has been analysed and the models are 'wrong' equally as often on the high and low side which means that, once aggregated, the statistics presented are unlikely to be skewed (i.e. the models are equally as likely to predict 70 filled posts for a location with 50 filled posts as they are to predict 30 filled posts).

Table 2. Proportion of location level estimates within predefined distance of actual ASC-WDS value, 2013 to 2024

Source: Skills for Care estimates

Estimation type	Metric	Care homes with nursing	Care homes without nursing	Non-residential
Imputation (forward extrapolation)	% within 10 filled posts of known value	90%	97%	85%
	% within 25% of known value	96%	92%	96%
	% within 10 filled posts or 25% of known value	97%	98%	97%
Care home machine learning model	% within 10 filled posts of known value	53%*	78%	n/a
	% within 25% of known value	87%	83%	n/a
	% within 10 filled posts or 25% of known value	88%	95%	n/a
Regression model (non-residential only)	% within 10 filled posts of known value	n/a	n/a	54%
	% within 25% of known value	n/a	n/a	72%
	% within 10 filled posts or 25% of known value	n/a	n/a	82%
Non-residential machine learning model (with dormancy)	% within 10 filled posts of known value	n/a	n/a	38%
	% within 25% of known value	n/a	n/a	71%
	% within 10 filled posts or 25% of known value	n/a	n/a	75%
Non-residential machine learning model (without dormancy)	% within 10 filled posts of known value	n/a	n/a	28%
	% within 25% of known value	n/a	n/a	62%
	% within 10 filled posts or 25% of known value	n/a	n/a	65%

* The percer19.3(*)]TJETQq0.000008871 0 595.32 8/F1 12 d TJETQq0.000008871 0 595.32 847percer1

5. Case Studies

The following case studies use example data and are intended to show how each step of the method would be applied to a location in practice. The first case study describes a location that has submitted data to ASC-WDS, whereas the second case study describes a location that has never submitted data.

5.1.

we can only see the 2021 data, and not their 2020 submission. Therefore, we filter out their data and use the value from the regression model to replace the very high data in ASC-WDS.

Location A doesn't update their data in 2022 and so the 900 value from 2021 is carried forward into our 2022 data. It is again replaced by a modelled value because it is so much higher than would be expected for a 20-bed care home.

In 2023, Location A corrects their data back to 90 filled posts and as in 2020, we use this value for our estimates.

Finally in 2024, Location A doesn't update their data again, meaning that the value of 90 is carried forwards from the previous year. This is again used for our estimates, see Table 4.

Table 4. Location A's estimates using the old methods over the past six years
Source: Example data

Year	ASC-WDS filled posts - raw	CQC number of beds	Estimated filled posts – old estimates
2019		20	25

New estimation method

Our new method takes several steps to smooth out and improve these estimates. The first step in our new methods is to remove repeated values across different snapshots over time (deduplication), see Table 5. The data in 2022 and 2024 were not updated submissions, so these values are removed. The data for 2020, 2021 and 2023 were all updated submissions, so they are retained in the deduplication stage. There was no data available in 2019, so this row remains blank.

Table 5. Location A's ASC-WDS submissions deduplicated over the past six years
Source: Example data

Year	ASC-WDS raw	CQC number of beds	Deduplication
2019		20	
2020	90	20	90
2021	900	20	

Finally, after the data has been deduplicated and filtered, we impute the missing values based on the data from other points in time for that location. We use a rolling average of the rate of change of the filled posts to bed ratio (for locations that have updated ASC-WDS data) to extrapolate the most recent known value forwards in time and the e

Chart 2. Location A's ASC-WDS submissions and estimates using the old and new methods over the past six years
Source: Example data

5.2. Case study 2: A care home without ASC-WDS data

Table 9. Location B’s estimates using the old methods over the past six years

Source: Example data

Year	ASC-WDS raw	CQC number of beds	Old estimates
2019		20	55
2020		20	51

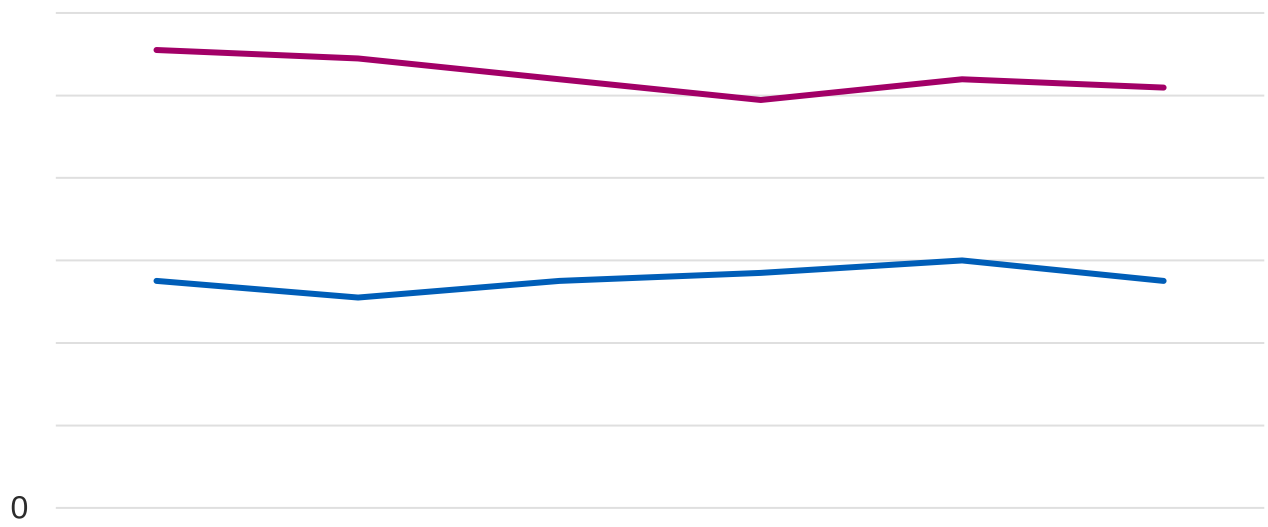
Table 11. Location C’s estimates using the old methods over the past six years
Source: Example data

Year	ASC-WDS raw	Old estimates	PIR raw
2019		55	
2020		51	100
2021		55	100
2022		57	90
2023		60	95
2024		55	95

PIR data was first collected in 2020, so we begin by de MC q184.58 655.3 84.624 15.84 le data

We can then use that to create a more accurate estimate than if we had used the overall average.

Chart 3. Location C's estimates using the new and old methods over the past six years
Source: Example data



6. Changes to our historical estimates

As these new estimates are different to our previously published estimates, we have undertaken a thorough analysis of these changes.

6.1 Differences between our new care home estimates and previous annual figures

Chart 4

Another reason for the difference is that the new process identifies dual registrations in the CQC Locations API and removes them (5,577 posts). Dual registered locations are locations registered with more than one provider and are included twice, with two different location IDs, in the CQC API. These duplicates were not identified by the old process.

Table 14. Difference between new and old independent CQC-regulated care home estimates by source and service type, March 2022
Source: Skills for Care estimates

New estimate source	Old estimate source	Number of locations	Difference: New estimate minus previous estimate	Average difference per location
Current ASC-WDS data	ASC-WDS data	1203	253	0.2
	Regression model	95	951	10.0
Machine learning model	ASC-WDS data	86	230	2.7
	Regression model	4274	-2072	-0.5
Imputed ASC-WDS data	ASC-WDS data	4656	-626	-0.1
	Regression model	4290	-5578	-1.3
CQC dual registrations		63	-5577	-88.5
Total difference		**14667**	**-13541**	**-0.9**

The largest difference in our estimates can be seen to be in 2020 (see Chart 4). This was around and after the pandemic when we started producing a monthly estimate for how filled posts were changing within the year.

From 2021 onwards we started to utilise these monthly figures to assist with our annual estimates and the differences since that point were smaller (it allowed us to do some of the imputation steps present in the new process). One of the benefits of our new automated process

Includes the impact of locations opening and closing
Uses precise update dates, rather than comparing 'before and after March 2024'.

Conclusion

Key Strengths

We are now able to:

- Provide monthly filled posts estimates for the number of filled posts at care homes in the CQC-regulated independent sector
- Provide monthly filled posts estimates for the number of filled posts at non-residential establishments in the CQC-regulated independent sector
- Provide monthly estimates of all locations in the CQC-regulated independent sector, not just those locations in ASC-WDS
- Provide stable trends over the history of our dataset
- Provide more accurate estimates of non-residential establishments compared to our previous model
- Examine our data over time, not as isolated years

We will soon be able to:

- Provide monthly filled posts estimates at regional, ICB, aE

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