

Annex B – overall assessment methodology

NHS England’s annual performance assessment of CCGs 2017/18

1. The CCG IAF comprises 51 indicators selected to track and assess variation across policy areas covering performance, delivery, outcomes, finance and leadership. This year, assessments have been derived using an algorithmic approach informed by statistical best practice; NHS England’s executives have applied operational judgement to determine the thresholds that place CCGs into one of four performance categories overall.

Step 1: indicator selection

2. A number of the indicators were included in the 2017/18 IAF on the basis that they were of high policy importance, but with a recognition that further development of data flows and indicator methodologies may be required during the year. However, by the end of the year, there was just one indicator that was excluded as there is no data available for the measure: mental health crisis.

Step 2: indicator banding

3. For each of the 207 CCGs, the remaining indicator values are calculated. For each indicator, the distance from a set point is calculated. This set point is either a national standard, where one exists for the indicator (for example in the NHS Constitution); or, where there is no standard, typically the CCG’s value is compared to the national average value.
4. Indicator values are converted to standardised scores (‘z-scores’), which allows us to assess each CCG’s deviation from expected values on a common basis. CCGs with outlying values (good and bad) can then be identified in a consistent way. This method is widely accepted as best practice in the derivation of assessment ratings, and is adopted elsewhere in NHS England and by the CQC, among others.¹
5. Each indicator value for each CCG is assigned to a band, typically three bands of 0 (worst), 2 (best) or 1 (in between).²

Step 3: weighting

6. Application of weightings allows the relatively greater importance of certain components (i.e. indicators) of the IAF to be recognised and for them to be given greater prominence in the rating calculation.
7. Weightings have been determined by NHS England, in consultation with operational and finance leads from across the organisation, and signal the

¹ Spiegelhalter et al. (2012) *Statistical Methods for healthcare regulation: rating, screening and surveillance*

² For a small number of indicators, more than 3 score levels are available, for example, the leadership indicator has four bands of assessment.

significance we place on good leadership and financial management to the commissioner system:

- Performance and outcomes measures: 50%;
- Quality of leadership: 25%; and,
- Finance management: 25%

8. These weightings are applied to the individual indicator bandings for each CCG to derive an overall weighted average score (out of 2).

Figure 1: Worked example

Anytown CCG has:

- Quality of leadership rating of “Green” (equivalent to a banded score of 1.33)
- Finance management rating of “Green” (equivalent to banded score of 2)
- For the remaining 48 indicators, the total score is 49.5.
- These scores are divided through by their denominator and weighted to produce an overall domain weighted score:

$$\left(\frac{1.33}{1}\right) \times 25\% + \left(\frac{2}{1}\right) \times 25\% + \left(\frac{49.5}{48}\right) \times 50\% = 1.35$$

Step 4: setting of rating thresholds

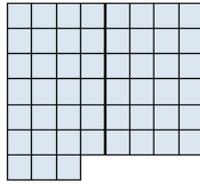
9. Each CCG’s weighted score out of 2 is plotted in ascending order to show the relative distribution across CCGs. Scoring thresholds can then be set in order to assign CCGs to one of the four overall assessment categories.
10. If a CCG is performing relatively well overall, their weighted score would be expected to be greater than 1. If every indicator value for every CCG were within a mid-range of values, not significantly different from its set reference point, each indicator for that CCG would be scored as 1, resulting in an average (mean) weighted score of 1. This therefore represents an intuitive point around which to draw the line between ‘good’ and ‘requires improvement’.
11. In examining the 2017/18 scoring distribution, there was a natural break at 1.45, and a perceptible change in the slope of the scores above this point. This therefore had face validity as a threshold and was selected as the break point between ‘good’ and ‘outstanding’.
12. NHS England’s executives have then applied operational judgement to determine the thresholds that place CCGs into the ‘inadequate’. A CCG is rated as ‘inadequate’ if it has been rated red in both quality of leadership and financial management.
13. This model is also shown visually below:

Deriving the CCG IAF assessment ratings

Step 1:

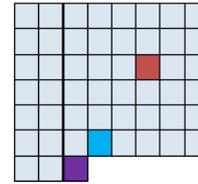
Indicators selected and calculated

There are 51 indicators in the 2017/18 CCG IAF...



...of which, 50 are included in the end of year rating

(1 indicator, **Mental health crisis team provision**, is excluded because data are not yet available)

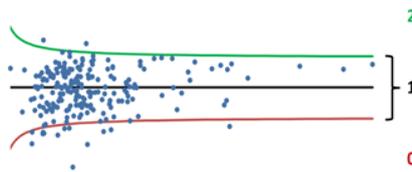


Values are derived for each CCG for each indicator. There is 1 indicator in the **Finance** domain and 1 for **Quality of Leadership**.

Step 2:

Indicators banded

Measure of deviation ("z-score") calculated for each CCG value. Outlying CCGs assigned to bands with scores of 0 (worst) to 2 (best).



The process is repeated for all 50 available indicators (example scores shown for **Anytown CCG**).

1	1	1	1	0	1	0	2
1	1	2	1	1	1	0	1
1	1	1	2	1	1	1	1
1	1	1	2	1	2	1	1
0	1	1	1	1	2	1	1
1	1	1	2	0	2	0	1
2	0	1					

Step 3:

Weights applied, average score calculated

Weightings set to:

- Finance: 25%
- Leadership: 25%
- The rest: 50%

Bandings for each domain are summed and divided by the count of indicators in that domain, then multiplied by the relevant weighting.

Worked example for Anytown CCG

Overall score calculated for CCG as sum of:

- [Finance] 25% * (2 / 1 indicator)
- + [Leadership] 25% * (1.333 / 1 indicator)
- + [The rest] 50% * (49.5 / 48 indicators)

= score of 1.35
(out of a possible 2)

Step 4:

Scores plotted and rating thresholds set

The distribution of average scores (out of 2) is plotted for all 207 CCGs. The threshold between "Requires Improvement" and "Good" is then set at the mid-point of 1; for "Outstanding" it is set at a natural break at the upper end of the distribution and for "Inadequate" an auto-rule is applied to include all CCGs whose Finance and Leadership ratings are both Red. In the example shown, there is a step change at 1.45 which forms the lower threshold for "Outstanding".

