Publications Gateway reference 07006

CCG IAF Methodology Manual

Purpose

To summarise the methods used in the production of indicators and ratings in the CCG IAF.

Introduction

The CCG IAF

The CCG Improvement and Assessment Framework (CCG IAF) provides a focus on assisting improvement alongside the statutory assessment function of NHS England. It aligns with NHS England's Mandate and planning guidance, with the aim of unlocking change and improvement in a number of key areas. This approach aims to reach beyond CCGs, enabling local health systems and communities to assess their own progress from ratings published online.

The Framework includes a set of 60 indicators, and at the end of the financial year, there is a process to derive an overall year end assessment for each CCG. A high level summary of the process can be found in Annex A.

Indicators

The list of indicators used is included the table in Annex B. Further detail about the indicators is in the Technical Annex, which is published in the section "Framework documents – Technical Annex" which is available here.

When choosing an indicator, NHS Digital's 'Criteria and considerations used to determine a quality indicator' was used as a guide.

Other things considered when selecting indicators were:

Time period

The aim was to provide denominators large enough to accurately identify as statistically significant material differences in performance.

For example, for an indicator with an average CCG proportion of 0.4 (40 per cent) based on an average of about 400 individuals per year, the standard error of a typical CCG's value based on three months' data was estimated as $\sqrt{((0.4 \times (1-0.4))/100)} = 0.048$, which would allow a difference of about 10 percentage points

from a reference indicator value (e.g. a standard) to be identified as statistically significant. If opinion was that a differences of five percentage points from standard was the minimum material difference and there was a need to identify such differences, then indicator values based on a quarter's data would not meet the need, as many CCGs would have values which were materially but not statistically significantly different from the standard. In such circumstances, use of 12 months' rather than three months' data was considered, as this would halve the estimated standard error, and allow such differences to be identified. If quarterly results were important, then use of a rolling twelve months' data recalculated every three months was considered.

Frequency

Once the required time period has been identified, the frequency was chosen to meet business needs, with the use of rolling data periods where needed.

Timeliness

The most recent available data was used, for preference matching with the formal time period of the CCG IAF. Typically, therefore, 2016/17, quarter four 2016/17, or March 2017 data was used for the end year 2016/17 CCG IAF, if available.

Missing data

Indicators were only used in the assessment if values for the majority of CCGs were available or if the indicator data source was published. Those indicators which were excluded on this basis are highlighted in Annex B.

Extreme values

The methodology for treating extreme values was robust, noting they needed to be excluded from any over-dispersion calculations (see section below on banding).

Standardisation and risk adjustment

Where needed, indicators were standardised or risk adjusted to provide a fair assessment of CCGs.

Assurance

The NHS Digital Indicator Assurance Service was used where assurance was needed.

Indicator banding

Purpose

To describe how scores were produced, using indicators, for each CCG.

The general approach and principles are set out below. Annex B shows, for each indicator, the specific approach used. **All scores were calculated on a 0 (bad) to 2 (good) scale**.

Measures of deviation

Where there was an agreed national standard, target, ambition or trajectory (table 1), the deviations which were scored were measured from the standard, target, ambition or trajectory value. Otherwise the deviations were measured from the England mean value.

Transformation and z scores

z scores were calculated for most indicators where this was possible, using transformation where necessary to stabilise the variance.

- For proportions (equivalently percentages) the arcsin√ transformation was used

  $z = 2\sqrt{n}\,(\arcsin\sqrt{r/n} - \arcsin\sqrt{p})$

Where the observed proportion had numerator *r* and denominator *n*, and *p* was the England mean proportion, or value of standard. The standard error in this case was

$$s = 1 \Big/ (2\sqrt{n})$$

- For indirectly standardised rates the √ transformation was used

  $z = 2(\sqrt{O} - \sqrt{E})$

Where *O* was the observed count and *E* was the expected count, if the England mean rate was applied. The standard error in this case was

$$s = 1 \Big/ (2\sqrt{E})$$

- Otherwise where available the value of the standard error *s* was used, or an estimate *s = (ucl-lcl)/3.92* where *ucl* or *lcl* were the upper and lower 95 per cent confidence limits was used; z was the deviation from the England mean or standard, divided by *s*.

Over-dispersion

For most indicators where *z* scores were used, over-dispersion corrections were applied. The calculation and application of the over dispersion parameter used the method described in Spiegelhalter, D.J (2005) [Funnel plots for comparing institutional performance. *Statistics in Medicine* **24**:1185-1202]. A random effects model was used, whereby excess variation in the CCG indicator values was assumed to be due to shortcomings in the risk adjustment processes. In calculating the over-dispersion parameterτ^2, 10 per cent of the CCG data values were winsorised (their values reset to the 10th or 90th percentile values) at each end of the distribution. *z* scores were then rescaled by multiplying by $\sqrt{s^2/(s^2+\tau^2)}$ where *s* was the standard error of the data point.

Over-dispersion corrections were not applied to indicators where there were agreed standards or targets (listed in the table below) as CCGs were expected to meet these irrespective of variation which might form part of a risk adjustment process. Corrections were applied, however, where there were national ambitions or trajectories which applied primarily at the national level, rather than being expected to be met by each CCG regardless.

Scores and thresholds

Where z-scores were available, they were converted to scores as follows:

- If $z < -1.96$, score 0
- If $-1.96 \leq z < 1.96$, score 1
- If $z \geq 1.96$, score 2.

In this case a large positive z corresponds to a “good” indicator value – the scale was reversed where necessary so that a score of 2 was always the “best”.

Where agreed standards (or targets, ambitions or trajectories) had been used in constructing the *z* scores, an alternative scoring system was used. The indicators affected are listed in the table below. Changes (by exception) were agreed between the relevant Clinical Panel and NHS England.

- If $z < -1.96$, score 0
- If $-1.96 \leq z < 0$, score 0.75
- If $0 \leq z < 1.96$, score 1.25
- If $z \geq 1.96$, score 2.

Again, the scale was reversed if needed so that a score of two was best.

Table 1: Indicators with standards, trajectories, targets or ambitions

Indicator reference	Name	Standard, trajectory, target and ambition values
Standards		
122b	Cancer – 62 days	85% (standard)
123a	MH – IAPT recovery	50% (standard)
123b	MH – EIP 2 weeks	50% (standard)
124a	LD – inpatients	TCP* specific (trajectory)
126a	Dementia – diagnosis	66.7% (standard)
127c	A+E – 4 hour	95% (standard)
129a	18 weeks	92% (standard)
107a	AMR – prescribing	1.161 or bespoke (target)
107b	AMR – broad spectrum	10% or bespoke (target)
Others		
122c	Cancer - survival	70.4% (trajectory)
125a	Maternity – neonatal mortality and still births	0.67% (ambition)

*TCP Targets apply to Transforming Care Partnerships, achievement against them is attributed to each CCG in the TCP.

Exceptions

Other types of indicator were scored directly:

- Red Amber Green (RAG) ratings were scored Red=0, Amber=1, Green=2 [or Red=0, Amber=0.67, Green=1.33, Green star=2 where there was a four point scale]
- Y/N ratings (Y good) were scored Y=2, N=0
- A direct relationship with good/bad was used where possible – e.g. a percentage based on a score of 0-15 where below 10 was bad would have 0-66.6 scored 0, 66.7 to 83.3 scored 1, above that scored 2.
- Otherwise quartiles or deciles were used - lowest scored 0, the highest scored 2, others 1.

For a small number of indicators, over-dispersion corrections were not applied due to the data needed to calculate them not being available.

Missing data

Where missing or seriously incomplete data represented a failing on the part of the CCG (for example, the failure to encourage adequate participation in the diabetes clinical audit), such data were scored as zero. Otherwise they were scored as one.

Extreme values

Extreme values were checked to ensure they were not errors. If they were, they were treated as missing (see above). If not, it was noted the methods are robust against the presence of extreme values, except if over-dispersion corrections were used as part of a z-scoring process for the indicator, such values were included in

the portion of the distribution which was winsorised prior to calculating the corrections.

Aggregation of scores

Purpose

To describe how the scores for each indicator arising from the “Indicator banding” stage are combined to give an overall score.

Weighting method

For each CCG, the overall score S was constructed as:

$$S = \sum_i w_i S_i$$

Where the CCG score for the *i*th indicator was S_i (a value between 0 and 2) and the weight given to the indicator was w_i.

Weights

The following weights were applied in the final rating calculation for 2016/17:

- Quality of leadership: 25 per cent; and,
- Finance management: 25 per cent (the assessment of financial plan is zero weighted to ensure focus on financial outturn)
- The remaining performance and outcomes measures: 50 per cent

Assessment ratings

Purpose

To describe the construction of the four category ratings.

Choice of thresholds – principles

The distribution of aggregated scores (0-2) by CCG informed the choice of thresholds. Furthermore, the following considerations were taken into account:

- As there were four ordered rating categories, three thresholds were needed to distinguish them.
- Where possible, natural breaks in the distribution were used as thresholds.
- The differences between thresholds were chosen where possible to be meaningful – so two CCGs between which there were no practically meaningful differences in the individual indicators fell either in the same (preferably) or in adjacent rating categories.
- At least some CCGs fell into each category
- Unless there were compelling reasons otherwise, it was expected there would be more CCGs in the middle two categories than in the extreme categories.

If a CCG was performing relatively well overall, their weighted score would be expected to be greater than one. 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 one, resulting in an average (mean) weighted score of one. This was therefore selected as an appropriate threshold between the two middle categories ‘good’ and ‘requires improvement’.

In examining the 2016/17 scoring distribution, a natural break was identified at 1.45. This was therefore selected as the threshold between the top and second categories.

CCGs were rated in the bottom category if they were rated “red” in relation to both quality of leadership and financial management.

Category names

The following labels are used for the four categories:

- Outstanding
- Good
- Requires improvement
- Inadequate

Presentation and Visualisation

Purpose

To describe how and where the indicator set is presented and visualised, and the processes governing its release.

MyNHS

The indicator set, including the end-of-year ratings is published on MyNHS

The indicators are presented by theme (better health, better care, sustainability, well-led) and area. The published CCG IAF is refreshed quarterly, although not all individual indicators are updated, and the model is updated annually.

Data Tool

NHS England and CCGs have access to the detailed indicators via the CCG IAF dashboard.

Underlying data

Most indicators were formed by secondary analysis of already published data. The CCG IAF is not intended as a vehicle for first publication of data. Underlying data values are however released on NHS England’s website.

Disclosure control

Where, as is the case for most indicators, they were formed from secondary analysis of already published data, issues of disclosure control did not arise. Where new primary data were being published, these complied with the NHS Anonymisation Standard.

Revisions

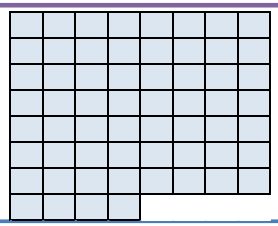
Where updated indicator values become available, the indicators will be reissued as part of the next regular quarterly release. In the event of significant errors coming to light between quarterly issues which are material at a national level and which go beyond the level of corrections normally expected from quarter to quarter, consideration will be given to issuing a special revision. Advice will be sought from the NHS England Head of Profession for Statistics.

Annex A: overview of the CCG IAF ratings production process

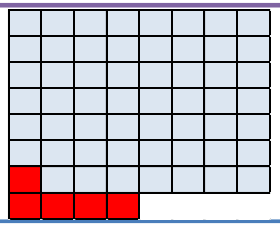
Deriving the 2016/17 CCG IAF assessment ratings

Step 1: Indicators selected

60 indicators in the IAF…



…of which, 55 included in the end-year rating calculation



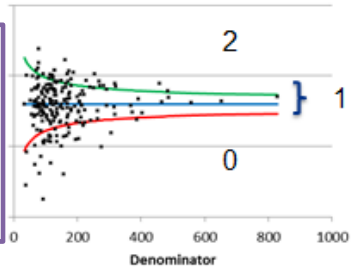
5 indicators excluded due to lack of data availability or completeness:

- End of life care (% of deaths in hospital)
- Ambulance waiting times
- 2x RightCare indicators
- 7 Day Services

Step 2: Indicators banded

Indicator values derived for each CCG

Measure of deviation calculated ("z-score") for each CCG value. Outlying CCGs assigned to bands with scores of 0 (worst), 2 (best) and 1 (the rest).



Process repeated for all 55 indicators

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

Step 3: Weights applied, average score calculated

Indicator weightings set:
- Quality of leadership: 25%
- Finance in-year: 25%
- The rest combined: 50%

*Note: finance plan is given zero weight

<u>Worked example for Anytown CCG above</u>

Average score calculated for CCG as sum of:
[Leadership] 25% * 0
+ [Finance] 25% * (1/1)
+ [The rest] 50% * (56/52)
= **0.79** (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 209 CCGs. The threshold between requires improvement and good is set at 1, and the outer bounds for the outstanding and inadequate categories are set by eye-balling the distribution to identify any natural breaks. NHS England executives have applied judgement to determine appropriate thresholds between categories.

In the worked example for Anytown CCG, 0.79 equates to requires improvement.

Annex B: Indicator Specification

Indicator Description						Outlier Calculation					Banding						
Ref	Indicator	Time period	Nature of indicator	What is good?	Include in assessment?	Deviation from	Transformation	Z scored?	Winsorisation level	Over dispersion correction	Score 0 if	Score 0.67	Score 0.75	Score 1	Score 1.25	Score 1.33	Score 2 if
101a	Maternal smoking at delivery	quarter	Proportion	Low	Yes	England mean	arcsinv	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
102a	Percentage of children aged 10-11 classified as overweight or obese	3 year	Proportion	Low	Yes	England mean	arcsinv	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
103a	Diabetes patients that have achieved all the NICE-recommended treatment targets	year	Proportion	High	Yes	England mean	arcsinv	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
103b	People with diabetes diagnosed less than a year who attend a structured education course	year	Proportion	High	Yes	England mean	arcsinv	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
104a	Injuries from falls in people aged 65 and over	year	Rate per popn over 65	Low	Yes	England mean	√	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
105a	Utilisation of the NHS e-referral service to enable choice at first routine elective referral	month	Rate per Gp referrals	High	Yes			No			p<0.5			0.5≤p<0.8			p≥0.8
105b	Personal health budgets	quarter	Rate per popn	High	Yes	Tracjectory		No			>50%from tracjectory			≤50%from trajectory>10%			≤10% from trajectory
105c	Percentage of deaths which take place in hospital				No (placeholder for end of life care in 2016/17, new indicator being introduced in 2017/18)												
105d	People with a long-term condition feeling supported to manage their condition(s)	year	Proportion	High	Yes	England mean	arcsinv	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
106a	Inequality in unplanned hospitalisation for chronic ambulatory care sensitive conditions	rolling year	Slope (directly standardised)	Low	Yes	England mean		Yes			z≥1.96			-1.96≤z<1.96			z<-1.96
106b	Inequality in emergency admissions for urgent care sensitive conditions	rolling year	Slope (directly standardised)	Low	Yes	England mean		Yes			z≥1.96			-1.96≤z<1.96			z<-1.96
107a	Anti-microbial resistance: Appropriate prescribing of antibiotics in primary care	rolling year	Rate per STAR PU	Low	Yes	Target. 1.161 or CCG bespoke	√	Yes			z≥1.96		Not achieving target and not scoring 0		Achieving target and not scoring 2		z<-1.96
107b	Anti-microbial resistance: Appropriate prescribing of broad spectrum antibiotics in primary care	rolling year	Proportion	Low	Yes	Target. 10% or CCG bespoke	arcsinv	Yes			z≥1.96		Not achieving target and not scoring 0		Achieving target and not scoring 2		z<-1.96
108a	Quality of life of carers	year	Score	High	Yes			No			lower quartile			middle two quartiles			upper quartile
121a	Provision of high quality care - Hospitals		Score	High	Yes			No			score<55.5%			55.5%≤score<66.6%			66.6%≤score
121b	Provision of high quality care - Primary Medical Services		Score	High	Yes			No			score<55.5%			55.5%≤score<66.6%			66.6%≤score
121c	Provision of high quality care - Adult Social Care		Score	High	Yes			No			score<55.5%			55.5%≤score<66.6%			66.6%≤score
122a	Cancers diagnosed at early stage	year	Proportion	High	Yes	England mean	arcsinv	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
122b	People with urgent GP referral having first definitive treatment for cancer within 62 days of referral	year	Proportion	High	Yes	Standard (85%)	arcsinv	Yes			z<-1.96		Not achieving standard and not scoring 0		Achieving standard and not scoring 2		z≥1.96
122c	One-year survival from all cancers	year	Proportion	High	Yes	Trajectory (70.4%)		Yes			z<-1.96		Not achieving trajectory and not scoring 0		Achieving trajectory and not scoring 2		z≥1.96
122d	Cancer patient experience	year	Proportion	High	Yes	England mean		Yes			z<-1.96			-1.96≤z<1.96			z≥1.96
123a	Improving Access to Psychological Therapies recovery rate	quarter	Proportion	High	Yes	Standard (50%)	arcsinv	Yes			z<-1.96		Not achieving standard and not scoring 0		Achieving standard and not scoring 2		z≥1.96
123b	People with first episode of psychosis starting treatment with a NICE-recommended package of care treated within 2 weeks of referral	rolling year	Proportion	High	Yes	Standard (50%)	arcsinv	Yes			z<-1.96		Not achieving standard and not scoring 0		Achieving standard and not scoring 2		z≥1.96
123c	Children and young people's mental health services transformation	quarter	Score	High	Yes			No			score<50%			50%≤score<90%			90%≤score
123d	Crisis care and liaison mental health services transformation	quarter	Score	High	Yes			No			score<50%			50%≤score<90%			90%≤score
123e	Out of area placements for acute mental health inpatient care - transformation	quarter	Score	High	Yes			No			score<50%			50%≤score<90%			90%≤score
124a	Reliance on specialist inpatient care for people with a learning disability and/or autism	quarter	Rate per popn	Low	Yes	TCP specific target	√	Yes			z≥1.96			-1.96≤z<1.96			z<-1.96
124b	Proportion of people with a learning disability on the GP register receiving an annual health check	year	Proportion	High	Yes	England mean	arcsinv	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
125a	Neonatal mortality and stillbirths	year	Proportion	Low	Yes	Trajectory (0.696%)	arcsinv	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96

Indicator Description						Outlier Calculation					Banding						
Ref	Indicator	Time period	Nature of indicator	What is good?	Include in assessment?	Deviation from	Transformation	Z scored?	Winsorisation level	Over dispersion correction	Score 0 if	Score 0.67	Score 0.75	Score 1	Score 1.25	Score 1.33	Score 2 if
125b	Women's experience of maternity services	year	Score	High	Yes	England mean		Yes			z<-1.96			-1.96≤z<1.96			z≥1.96
125c	Choices in maternity services	year	Score	High	Yes	England mean		Yes			z<-1.96			-1.96≤z<1.96			z≥1.96
126a	Estimated diagnosis rate for people with dementia	month snapshot	Rate per dem pop	High	Yes	Ambition (2/3)	√	Yes			z<-1.96		Not achieving ambition and not scoring 0		Achieving ambition and not scoring 2		z≥1.96
126b	Dementia care planning and post-diagnostic support	year	Proportion	High	Yes	England mean	arcsin√	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
127a	Achievement of milestones in the delivery of an integrated urgent care service	quarter	Score	High	Yes			No			score<4			4≤score<8			score=8
127b	Emergency admissions for urgent care sensitive conditions	quarter	Rate per popn (directly standardised)	Low	Yes	England mean	√	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
127c	Percentage of patients admitted, transferred or discharged from A&E within 4 hours	year	Proportion	High	Yes	Standard (95%)	arcsin√	Yes			z<-1.96		Not achieving standard and not scoring 0		Achieving standard and not scoring 2		z≥1.96
127d	Ambulance waits				No (data unavailable for pilot sites)												
127e	Delayed transfers of care attributable to the NHS per 100,000 population	month	Rate per popn	Low	Yes	England mean	√	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
127f	Population use of hospital beds following emergency admission	quarter	Rate per popn (indirectly standardised)	Low	Yes	England mean	√	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
128a	Management of long term conditions	quarter	Rate per popn (directly standardised)	Low	Yes	England mean	√	Yes	10%	random effects	z≥1.96			-1.96≤z<1.96			z<-1.96
128b	Patient experience of GP services	annual	Proportion	High	Yes	England mean	arcsin√	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
128c	Primary care access	quarter	Proportion	High	Yes			No			score<1/3			1/3≤score<2/3			score>2/3
128d	Primary care workforce	quarter snapshot	Rate	High	Yes	England mean	√	Yes	10%	random effects	z<-1.96			-1.96≤z<1.96			z≥1.96
129a	Patients waiting 18 weeks or less from referral to hospital treatment	month snapshot	Proportion	High	Yes	Standard (92%)	arcsin√	Yes			z<-1.96		Not achieving standard and not scoring 0		Achieving standard and not scoring 2		z≥1.96
130a	Achievement of clinical standards in the delivery of 7 day services				No (data source not published)												
131a	People eligible for standard NHS Continuing Healthcare	quarter	Rate per popn	Low	Yes			No					upper and lower decile		10th to 90th deciles		
141a	Financial plan	year	RAG	Green	Yes			No			Red			Amber			Green
141b	In year financial performance	quarter	RAG	Green	Yes			No			Red			Amber			Green
142a	Outcomes in areas with identified scope for improvement				No (data only available for 65 pilot sites)												
142b	Expenditure in areas with identified scope for improvement				No (data only available for 65 pilot sites)												
143a	Adoption of new models of care	quarter	Yes/No	Yes	Yes						No						Yes
144a	Local digital roadmap in place	quarter	Yes/No	Yes	Yes			No			No						Yes
144b	Digital interactions between primary and secondary care	quarter	Composite metric	High	Yes			No					lower quartile	middle two quartiles	upper quartile		
145a	Local strategic estates plan (SEP) in place	year	Yes/No	Yes	Yes			No			No						Yes
161a	Sustainability and Transformation Plan	year	RAG	Green	Yes			No			Red			Amber			Green
162a	Probity and corporate governance	quarter	3 point rating	Fully compliant	Yes			No			Not compliant			Partially compliant			Fully compliant
163a	Staff engagement index	year	Composite metric	High	Yes			No			score<3.75			3.75≤score<3.85			3.85≤score
163b	Progress against workforce race equality standard	year	Composite metric	Low	Yes			No					lower quartile	middle two quartiles	upper quartile		
164a	Effectiveness of working relationships in the local system	year	Score	High	Yes			No			score<60			60≤score<70			70≤score
165a	Quality of CCG leadership	quarter	RAGG*	Green (star)	Yes			No			Red	Amber				Green	Green Star