Web content, Oct 2022

Assessing services

We're developing a new approach to regulation. Until we implement this new assessment framework, we'll continue to use our current methods to monitor, assess and rate providers.

What will be different

- **Gathering evidence:** We'll make much more use of information, including people's experiences of care services. We'll gather evidence to support our judgements in a variety of ways and at different times – not just through on-site inspections. This means on-site inspections will support this activity, rather than being our primary way to collect evidence.
- **Frequency of assessments:** How often we assess will not be determined only by the previous rating and type of service. Evidence we collect or information we receive at any time can trigger an assessment.
- **Assessing quality:** We'll make judgements about quality more regularly, instead of only after an inspection as we do today. We'll use evidence from a variety of sources and look at any number of quality statements to do this. Our assessments will be more structured and transparent, using evidence categories as part of the assessment framework and giving a score for what we find. The way we make our decisions about ratings will be clearer and easier to understand.

We will refine and improve how this approach works as we start to test it and introduce it for providers.

Up-to-date, transparent assessments of quality

Our assessment framework gives us flexibility to:

- update the ratings for key questions and overall ratings when things change, based on more frequent assessment of evidence
- collect and review evidence in some categories more often than others. For example, we may collect evidence of people's experiences more often than evidence about processes
- be selective in which quality statements we look at – this could be one, several or all.

How often we assess services

The frequency of assessments will depend on the information we receive and the evidence we collect about a service, rather than just on the type of service and its previous rating. We use our approach to decide:

- where we focus our activity at a sector and individual provider level
- how often we carry out assessment activity
- what type of activity we use to gather evidence.

This allows us to collect evidence in both a planned and responsive way.

Planned activity

Each evidence category in the assessment framework has an initial schedule for ongoing assessment. This sets out the length of time before we need to collect evidence for that category in each service type. These are planned evidence collection activities.

This schedule is based on what we understand about quality in a particular type of service across the country. We may need to carry out planned evidence collection activity more frequently for a specific type of service, depending on:

- additional national priorities, for example in maternity services
- whether our view of risk in an individual service or area has changed.

Our current ambition is to update the information we hold on a service across all required evidence categories within a 2-year period. We won’t always collect and assess evidence at a fixed point during this 2-year period.

Responsive activity

As well as our planned evidence collection activity, we also receive regular pieces of information.

If we receive information that indicates an immediate risk, concern or change in quality, this could trigger action to collect evidence. For example, this could include information from:

- whistleblowing concerns
- safeguarding reports

- statutory notifications
- people using services, including information reported through give feedback on care

Sometimes, we will prioritise which services to assess according to where we think there might be a significant change in quality.

How we gather evidence

To decide whether to collect evidence on site or off site, we look at the required evidence categories for a quality statement and the type of service.

Examples of some evidence that we can collect entirely off site can include:

- data on outcomes of care, such as:
    - national clinical audits
    - patient reported outcome measures (PROMs)
    - infection control rates
- anonymised information from people's records
- interviews with staff and professionals who work in the service
- feedback and complaints
- evidence from people and their representatives about their experiences.

Some evidence can only be collected on site through an inspection, for example:

- observing the care environment
- talking to people about their experiences in some types of service
- understanding the culture and how staff interact with each other.

We will also work with other people and organisations to help us collect evidence, for example local Healthwatch groups and our Experts by Experience. They can help us reach out to people, families and carers and engage with communities whose voices are seldom heard. This means that we may not always need to physically enter a service to gather this evidence and update our ratings.

We will carry out site visits more frequently where:

- there is a greater risk of a poor or closed culture going undetected and it is the only way to gather people's experience of care
- it is the only way to ensure the right people and activities will be available to assess quality
- we have concerns about transparency and availability of evidence

- we have a statutory obligation to do so. For example, as a member of the National Preventative Mechanism we must visits places of detention regularly to prevent torture and other ill-treatment.

Our teams will also use the expertise of our Experts by Experience, specialist advisors and executive reviewers to inform our assessment activity. (Executive reviewers are colleagues who support on inspections of the well-led key question for NHS trusts). This ensures that our judgements maintain credibility. Assessment teams can get quick access to specialists to support them in:

- understanding which evidence to collect
- corroborating and analysing evidence
- interviewing key staff.

How we reach a rating

We'll continue to describe the quality of care using our 4 ratings: outstanding, good, requires improvement, or inadequate. In forming a view of quality, we use a scoring framework to enable us to make consistent judgements. Our scores will translate into one of the ratings for the key questions (safe, effective, caring, responsive, and well-led). Scores will also be the basis for our view of quality at an overall service level.

At first, we will only publish the ratings, but we also intend to publish the scores in future.

Using scoring as part of our assessments will:

- help us be clearer and more open about how we've reached a judgement on quality
- show if a service is close to another rating. For example, for a rating of good the score can show if it's nearing either outstanding or requires improvement
- help us to see if quality is moving up or down within a rating.

Our quality statements clearly describe the standards of care that people should expect.

To make things clearer, we'll set out the types of evidence we'll focus on in each required evidence category when we're assessing a quality statement.

To assess a particular quality statement, we will take into account the evidence we have in each of the required evidence categories. This will vary depending on the

type of service. For example, the evidence we will collect for GP practices will be different to what we'll have available to us in an assessment of a home care service.

Evidence could be information that we either:

- already have, for example from statutory notifications
- actively look for, for example from an on-site inspection

Depending on what we find, we give a score for each required evidence category:

4 = Evidence shows an exceptional standard of care

3 = Evidence shows a good standard of care

2 = Evidence shows shortfalls in the standard of care

1 = Evidence shows significant shortfalls in the standard of care

As we are moving away from assessing at a single point in time, in future we will likely assess different areas of the framework on an ongoing basis. This means we can update scores for different evidence categories at different times. This is because our national schedule guides how often we need to collect evidence and we'll want to do this more frequently in some services and for some categories than others.

Any changes in evidence category scores can then update the existing quality statement score. This can then have an impact on the rating.

We follow these stages to produce a key question and overall rating for a service:

1. Review evidence types within the required evidence categories for each quality statement.
2. Apply a score to each of these evidence categories.
3. Combine these required evidence category scores to give a score for the related quality statement.
4. Combine the quality statement scores to give a total score for the relevant key question.
5. This score generates a rating for each key question.
6. Aggregate the key question ratings to give the overall rating.

We weight all evidence categories and quality statements equally. Our operational teams will use their professional judgement when making decisions about quality. Our judgements go through quality assurance processes.

Example: how we reach a rating (GP practices)

To assess quality against a particular quality statement, operational colleagues specialising in GP practices will look at the required evidence categories. In this example, we are just looking at the infection prevention and control quality statement.

> "**Infection prevention and control:** We assess and manage the risk of infection. We detect and control the risk of it spreading and share any concerns with appropriate agencies promptly."

In general practice, the required categories for this are:

- people's experiences
- feedback from staff and leaders
- observation
- processes.

We would look at individual pieces of evidence under each category and based on the strength of what we find, give a score of 1 to 4.

For example, in the 'people's experience' evidence category, we may look at:

- patient surveys
- complaints and compliments

To gather evidence in the 'feedback from staff and leaders' and 'observation' categories, we might schedule:

- an inspection to look at the care environment
- a call to speak with staff at the GP practice.

We would then combine this new evidence with what we already hold on 'processes' to help us form a view of quality.

Example: combining evidence category scores to give a quality statement score

Evidence category	Score
People's experiences	3
Feedback from staff and leaders	2
Observation	3
Processes	3
Total score for the combined evidence categories	11

We calculate this as a percentage so that we have more detailed information at evidence category and quality statement level, and can share this. In time, this will support benchmarking information.

To calculate the percentage, we divide the total (in this case 11) by the maximum possible score. This maximum score is the number of required evidence categories multiplied by the highest score for each category, which is 4. In this case, the maximum score is 16. Here, it gives a percentage score for the quality statement of 69% (this is 11 divided by 16).

We convert this back to a score so it is easier to:

- understand
- combine with other quality statement scores to calculate the related key question score.

We use these thresholds to convert percentages to scores:

- 25 to 38% = 1
- 39 to 62% = 2
- 63 to 87% = 3
- over 87% = 4

In this case, the percentage score of 69% converts to a score of 3.

We then use this score to give us an updated view of quality at key question level. In this case it is for the safe key question:

Quality statement	**Score,**
Learning culture	2
Safe systems, pathways and transitions	3
Safeguarding	3
Involving people to manage risks	2
Safe environments	3
Infection prevention and control	3
Safe and effective staffing	2
Medicines optimisation	3
Total score for the safe key question	21

Again, we calculate a percentage score. We divide the total (in this case 21) by the maximum possible score. For the safe key question, this is 8 quality statements multiplied by the highest score for each statement, which is 4. So the maximum score is 32. Here, it gives a percentage score for the key question of 65.6% (this is 21 divided by 32).

At key question level we translate this percentage into a rating rather than a score, using these thresholds:

- 25 to 38% = inadequate
- 39 to 62% = requires improvement
- 63 to 87% = good
- over 87% = outstanding

Therefore, the rating for the safe key question in this case is good.