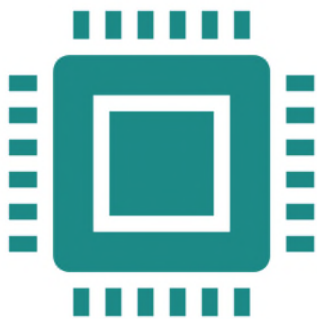


# Using machine learning in diagnostic services

A report with recommendations from CQC's regulatory sandbox



March 2020

# Contents

- Summary and key findings** **2**
- 1. Overview** **3**
- 2. The existing regulatory framework** **6**
- 3. Key challenges for deployment and assurance** **8**
- 4. What CQC will look for when inspecting and rating services** **9**
- 5. Recommendations** **10**
- 6. Conclusion** **12**
- 7. How we carried out this regulatory sandbox** **12**
- 8. What good looks like – detailed statement** **14**

# Summary and key findings

This report presents the findings from the Care Quality Commission's (CQC's) regulatory sandbox pilot. Regulatory sandboxing is a way of working proactively and collaboratively to understand new types of health and social care service, agree what good quality looks like, and develop our approach to regulation. We think this is particularly important for innovative and technology-enabled services, which are developing quickly, and where a response requires collaboration with other national bodies.

This sandbox round focused on the use of machine learning applications for diagnostic purposes in healthcare services. Part of this work involved building a consensus on what is needed to deliver high-quality care in services that use these applications. To do this, we worked with healthcare providers, technology suppliers, people who use services, clinicians, and other stakeholders. We have used the findings of this sandbox to identify and consider where we need to update our current regulatory methods, and what work we need to do to get this right, which will help us to regulate these services better.

## Key findings and recommendations from this sandbox

- Providers that use machine learning in diagnostic services need to have good governance in relation to the clinical, information, technical and human aspects of the application. We still need to develop our approach with providers to get this right.
- Most suppliers of machine learning applications in diagnostics will not need to register with CQC. Only those suppliers that deliver clinical activity themselves within the scope of a regulated activity need to register. Some parts of the market are developing quickly, and we anticipate having to register the first organisation that uses an autonomous machine learning system in routine clinical service before the end of 2020.
- To effectively regulate these few suppliers that become registered providers, and assure the public that their services are safe and effective, CQC will need other national bodies to develop technical standards and assess against them.
- There are two key gaps around the assurance of machine learning systems that national bodies need to address, particularly for autonomous systems: firstly there is a need for more guidance and infrastructure to support clinical validation of algorithms, both at the CE kitemarking stage and when implementing in a new site; secondly we need more clarity on how hospitals should implement machine learning devices within clinical pathways to ensure high-quality care.
- Technology suppliers need to clearly communicate what their products, solutions and devices do and how they perform. Suppliers do not always accurately state whether their products use machine learning. This makes it harder to implement devices safely and poses a risk to patients.

# 1. Overview

## 1.1 What is regulatory sandboxing?

A regulatory sandbox is a way of testing how best to regulate new types of services by working collaboratively to find out about them.

This is the second of three pilot rounds of sandboxing in 2019/20 where we have tested this approach. This work was supported by a £500,000 grant from the *Regulators' Pioneer Fund* launched by the Department of Business, Energy and Industrial Strategy (BEIS) and administered by Innovate UK. The fund enables UK regulators to develop an approach that enables innovation around emerging technologies, and to unlock the long-term economic opportunities identified in the government's Industrial Strategy.

This regulatory sandbox round focused on the use of machine learning applications in diagnostic services. We needed to understand what good quality care looks like in services that use these technologies, the best way to regulate them, and what CQC and other bodies can do to improve quality and support innovation. To do this, we worked with healthcare providers, technology suppliers, people who use these services, government partners, and other experts.

## 1.2 Machine learning and why it is important

Machine learning is a set of software algorithms and statistical models that computer systems use to perform a specific task, without using explicit instructions. Through this sandbox, we saw different devices and ways of using machine learning. Nearly all these devices were Convolutional neural networks that use supervised learning; but some suppliers have used other statistical techniques to develop their devices – all of which typically require large amounts of carefully curated and labelled training data.

This approach is different from other types of software where coding is done intentionally and transparently, based on what developers know. It introduces new issues, such as uncertainties about how the software is working, and differences in how performance is tested and improved. However, in health care the approach can be applied to a wide range of diagnostic-related data and could contribute to major improvements in outcomes for patients, through faster and potentially more accurate identification of disease or clinical risks to people's health. The government and many leaders in technology and health care have identified machine learning and other artificial intelligence techniques as a breakthrough technology, which is likely to bring big benefits to services and the people that use them. The Secretary of State for Health and Social Care has made an historic funding commitment of £250m to support the UK to become a global leader in artificial intelligence in our sector.<sup>a</sup>

<sup>a</sup> [Announcement of the National Artificial Intelligence Lab to improve the health and lives of patients \(2019\)](#)

Furthermore, the power of these new software methods allows the automation of some tasks that could previously only be performed by highly-trained professionals, such as radiologists. This is an important issue for regulation as regulatory responsibilities and methods have until this point assumed that trained clinicians would deliver these clinical activities.

### 1.3 Delivering machine learning applications in diagnostics and screening

There are many different types of products and approaches to applying machine learning in diagnostics. Different categories have different risks and requirements.

The U.S. Food and Drug Administration has set out a categorisation based on whether the software treats or diagnoses directly, whether it drives clinical management or informs clinical management, and the level of clinical risk involved in the pathway of care.<sup>b</sup>

We built on this framework for our sandbox, and defined the level of autonomy in three categories:

- Software that operates independently or without direct clinical supervision for at least some patients.
- Software that works independently but the clinician retains responsibility for the patient's management, and reviews results before taking action. Within this category, we include software that shapes pathways or conducts 'primary' reads, and software that shadows or checks what a clinician has done with a 'second' read.
- Software that provides information or support to clinicians to do their work, which we define as computer-aided detection (CAD). We see this as being similar to the older CAD software that has not relied on machine learning techniques such as Convolutional neural networks in the past.

We learned that the safety of a machine learning application is influenced not just by the clinical risks for each patient, but also by the clinical breadth of the application. We have seen that broad machine learning applications, or broad pathways, introduce human factors that increase risk. It is harder for both technology suppliers and clinicians to understand what these systems do and don't do for each patient. Nevertheless, it is important not to rule out broad based applications, as they are the most likely to help release capacity in clinical teams to deal with the increasingly unmanageable need for diagnostics in our healthcare system. We must therefore pay more attention to how deployments are managed in a clinical workflow, how clinical teams are trained, and how manufacturers work with their clients where the tools they are using seek to address a broader range of clinical use-cases.

<sup>b</sup> U.S. Food and Drug Administration, [Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning \(AI/ML\) – Based Software as a Medical Device \(SaMD\)](#) (2019).

## Examples of existing clinical machine learning applications in diagnostics

*Darker purple boxes indicate that more care is needed when using these solutions in live clinical services.*

### Level and scope of clinical risk

|  | Narrow: low clinical risk  | Narrow: high risk pathway   | Broad: high risk pathway   |
|--|--|---|--|
| <b>Level of autonomy</b>   |  | IDx-DR – diabetic retinopathy screening (FDA approved for direct to consumer activity)  | Plans for Behold.AI / Dartford and Gravesham NHS Foundation Trust to rule out high normal chest X-rays<br><br>Plans for Oxipit to perform chest x-ray reporting<br><br>Qure’s delivery of AI for TB and other lung disease in Indian systems |
| <b>Autonomous clinical software</b>  |  |   |  |
| <b>Clinical software working independently, under a clinician’s supervision for each patient</b> | Zebra Medical’s fatty liver detection algorithm  | Rapid diagnostic analysis of whether a stroke is ischaemic or haemorrhagic, where a clinician retains full responsibility prior to starting treatment<br><br>Kheiron / EMRAD proposals for Breast Screening pathway | Oxipit and Behold.AI in their current pilots with NHS trusts, which prioritise patients for radiologist review<br><br>ResApp, using cough as a biomarker to differentially diagnose lung-based presentations                                 |
| <b>Computer aided detection (CAD)<sup>c</sup></b>  | A lot of software is in this category, and it has been for several decades. However, there are now also machine learning-based products on the market, such as Aidence and Veye’s lung nodule check, ZebraMed’s heart calcification detection algorithm, and Healthy.IO’s AI based urine dipstick analysis for chronic kidney disease screening. |   |  |

<sup>c</sup> By CAD, we mean software that is comparable in how it operates as part of a clinical pathway to the more traditional forms of computer-aided detection and older forms of computer-aided diagnosis that are not using Convolutional neural networks or similar machine learning approaches to computer vision problems.

## 2. The existing regulatory framework

**The Medicines and Healthcare products Regulatory Agency (MHRA)** regulates medical devices across the UK. Software intended to provide diagnostic or therapeutic information is regulated as a medical device. MHRA's regulatory duties are currently set out in the Medical Device Regulations 2002 and amendments. The regulations are changing, with major amendments coming into force on 26 May 2020, which will reclassify many software devices into higher risk categories. Medical devices require a clinical evaluation; in many cases, this may require a clinical investigation. Medium and high-risk devices need to use a Notified Body. MHRA designates and audits notified bodies in the UK. Harmonised standards (European adoptions of ISO standards), may help to show conformity with the general safety and performance requirements of the device regulation.

**The Care Quality Commission (CQC)** regulates providers that carry on one or more of the regulated activities set out in the [Health and Social Care Act 2008 \(Regulated Activities\) Regulations 2014](#). Suppliers of machine learning applications ONLY need to register when they are performing these tasks independently from clinicians (rather than to support clinicians with additional data or insight). We anticipate that the first machine learning applications that fall into our scope of regulation will be those that are analysing and reporting on X-ray, CT and MRI. However, where a healthcare provider is using all types of machine learning applications that are important to delivering regulated activity, we need to understand how well they are working for patients. If necessary, CQC has the powers to review key third-party technology suppliers responsible for an activity ancillary to regulated activity.<sup>d</sup> We have not taken such action with CAD software, and we do not anticipate that our approach to machine learning software that supports clinicians in this way will differ substantially in that respect.

**NHSX** commissions relevant guidance from **NHS Digital** and has an important policy role in setting out and developing the regulatory infrastructure that we need to get this right. The main NHS Digital standards are: [DCB0160: Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems](#), [DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems](#) and [Information Governance and Technology Guidance](#). However, there are also several others around [identity](#), [information governance](#), and [interoperability](#), which we may want to expect of suppliers and services that interact with them. NHSX, in collaboration with the Accelerated Access Collaborative, is also responsible for the delivery of the National AI lab and related AI award.<sup>e</sup>

d. "An activity which is ancillary to, or is carried on wholly or mainly in relation to a regulated activity, shall be treated as a regulated activity", (Health and Social Care Act 2008 (Regulated Activities) Regulations 2014).

e. <https://www.england.nhs.uk/aac/what-we-do/how-can-the-aac-help-me/ai-award>

**The ISO standards<sup>f</sup>** that underpin medical device regulation (by MHRA) and data handling are of a high quality and complement NHS Digital's clinical risk management standards. ISO13485, regarding quality management systems, helps give assurance that technology suppliers have good quality management systems and governance structures. ISO27001 gives guidance on handling and processing data. ISO 14971 covers risk management and ISO 62304 covers software life-cycle processes. **British Standards Institute (BSI)** is the UK National Standards Body and supports the development and publication of ISO standards.

**Public Health England (PHE)** provides quality standards and guidance for all population screening programmes. These set out the requirements for services providing screening, standards to assure the quality of care, and key performance indicators to monitor delivery at a population level.

**The National Screening Committee (NSC)** is responsible for advising the Secretary of State for Health and Social Care on which technologies are sufficiently well evidenced to be used within a population screening programme.

**The National Institute for Health and Care Excellence (NICE)** has published [evidence standards for digital health technologies](#), and is currently considering how Health Technology Appraisal and Evaluation would work in the context of AI applications (outside of a screening context).

**The Health Research Authority (HRA)** regulates and manages clinical research, which we expect most machine learning application developers and adopters to be engaged in.

**The Information Commissioner's Office (ICO)** is the UK's independent body set up to uphold information rights in the interest of the public.

**Medical royal colleges**, such as the Royal College of Radiologists, play a leading role in setting professional standards of practice, and setting medical education curricula.

**NHS Improvement** has a financial and safety focused regulatory role for providers of NHS services, and within that manages a national system and team for identifying and alerting providers to safety issues.

f. See the [list of standards in MHRA guidance](#) (page 37), which is under review with the transition to the Medical Device Regulations



## 3. Key challenges for deployment and assurance

### 3.1 Integration with clinical services

**The way in which machine learning devices are implemented within a service is crucial to the quality of care that they deliver.** There are currently significant challenges around the implementation of machine learning devices, which mean it is not always done well.

- Providers do not always understand what the device can and cannot do – for example, if it can identify a specific set of conditions, but will not pick up other abnormalities – which can lead to inappropriate use and unsafe care.
- It can be challenging to ensure that the clinicians who interact directly with machine learning systems understand how they work and how to use them.
- Interoperability challenges mean that it can be hard to integrate machine learning devices with existing PACS/RIS/electronic health record (EHR) systems.

There is a lack of guidance and oversight of how machine learning devices are implemented and how they integrate with clinical services. This is not included within MHRA's regulation and CQC does not have the expertise to make technical assessments. Further work is needed to address these gaps.

**Care services do not always have the capability they need to monitor how systems are performing.** Auditing performance after implementation allows services to detect and respond to problems as they emerge, which is crucial to ensuring the ongoing quality and safety of care. Providers need to build the capability to do this well. This can be supported by clear national standards that can improve the quality of care and give providers confidence to adopt these technologies safely.

### 3.2 Assessing how well machine learning devices perform across different settings and populations

**The performance of machine learning devices can vary between different settings, sites or populations.** This means that providers cannot rely on test results from other sites or populations to assure themselves that a device will deliver high-quality care in their service. In the absence of national validation processes, some providers are undertaking validation locally, which can lead to inappropriate testing, additional costs for providers, and delays to delivery.

This is also a gap in regulatory assurance. While MHRA assures that devices perform in line with manufacturers' claims, there is no technical assurance that they deliver high-quality care when implemented within a service.

**There is a lack of understanding of how machine learning devices perform across different population groups.** Bias occurs naturally in all Convolutional neural networks, so it is an important issue for machine learning in healthcare settings. However, we did not find any formal studies that quantify this effect in diagnostic imaging. A major barrier to understanding this issue better is that manufacturers currently lack access to data on ethnicity, gender, and other data that they would need to quantify how much of a problem this is. This may be leaving some patient groups at risk of unsafe care.

## 4. What CQC will look for when inspecting and rating services

### 4.1 Providers of health and care services that use machine learning for diagnostic purposes

At a provider level, this regulatory sandbox has found varying degrees of clinical oversight, governance, and support for innovators and teams that want to use these technologies. We hope that this report, along with the impetus from the Department of Health and Social Care and the AI lab, will encourage leaders in providers such as hospitals to take stock of the benefits and risks of these technologies and provide adequate support and oversight to their clinical and digital teams to use it in the right ways.

Where CQC inspection teams include machine learning software as part of their inspections of a service, they will be encouraged to look at whether:

1. The tools are accessible, user-friendly and do not frustrate patients and staff.
2. The tools comply with product regulation.
3. The clinical information from tools is given in a way that works for clinical teams and those requesting diagnostics. This will typically require integration with the requesting and reporting systems (for example, PACS and RIS).
4. Clinical teams and those interacting with the tool have a good understanding of what it does and doesn't do, and that training and information have been carefully considered.
5. Safety is managed well and follows NHS Digital's Clinical Risk Assessment guidance DCB0160 and DCB0129 – both in terms of planning for risk, and identifying and addressing issues and incidents. As part of this, we will expect to see that the accountabilities for clinical processes, clinical content, software updates and training are clear between parties, and that there is a working relationship between services, those procuring the solution, and technology suppliers.
6. The development of clinical pathways for these tools is governed effectively, testing happens before tools are put into use, and there is evidence of clinical input and regular audit.

## 4.2 Organisations delivering clinically autonomous machine learning applications in diagnostics

Most organisations delivering machine learning will not need to register with CQC. However, if CT, MRI or physiological measurement is analysed or reported on by a machine learning tool without the oversight of a clinician for at least some patients in a live clinical setting (for example, only high confidence diagnoses or high confidence normals), then it becomes regulated activity. This means that it is an offence to deliver this without registering with the Care Quality Commission.

We expect to register the first providers using machine learning software in this calendar year, on a case-by-case basis. To enable us to assess these providers against our key questions, we will apply our assessment framework for healthcare services and adapt the teleradiology provider service framework that we currently use to adjust the questions that we ask. We will use our consensus statements of what good looks like as the basis for this work. It will be some time before we have registered and inspected enough providers to be able to rate their performance for the public.

## 5. Recommendations

This regulatory sandbox brought together people who use care services, healthcare providers, technology suppliers, regulators (such as CQC and MHRA), and other organisations that shape the health and care landscape in England (such as NHSX). The learning from the early stages of the process had a much broader application than just CQC's regulation, and it showed that CQC working alone can only have limited impact in driving improvements in the quality of care in this area. One output of the sandbox is therefore a set of recommendations for other organisations to consider.

- 1. Providers should have confidence in using machine learning applications that act as 'CAD' software** (meaning they are assisting but not replacing clinicians). However, to ensure high-quality care, providers **must** procure them diligently, use NHS Digital's clinical safety standards or an equivalent quality management process during deployment, and check that commercial products that they want to use have their kitemarks as medical devices. National and standards setting bodies should help CQC, AAC and NHSX to encourage the use of these registered medical devices where they have benefits to people.

**2. Regulators, industry, evidence and standard-setting bodies, and the scientific community need to work together to address two key knowledge gaps:**

- the accuracy of machine learning applications can vary significantly between different settings and populations;<sup>9</sup> what validation is required and how should it be organised to ensure everyone gets high-quality care?
- how should autonomous AI systems be monitored and regulated after they receive market approval, to ensure that the care they provide is always of a high quality and to build public trust?

**3. Regulators and other standard-setting bodies need to provide more clarity to care providers and software developers to address three important guidance gaps on what they need to do to deliver high-quality care:**

- The way in which machine learning technologies are implemented within a service is crucial to the quality of care they provide. **NHSX** should continue to work with **CQC**, the **royal colleges**, and other relevant bodies to develop clear standards that help care providers understand how to adopt these technologies within their services in a way that delivers safe, high-quality care.
- There is a need for greater clarity on how data can be used when developing and implementing machine learning technologies. The **ICO** should produce some specific guidance to help manufacturers and care providers maximise the benefits of these technologies while complying with relevant legislation such as GDPR.
- **MHRA** should clarify when a change to a machine learning algorithm is significant enough to require new regulatory approvals.

**4. National bodies should address the significant barriers around interoperability.** Some of the most common patient records, PACS or RIS systems cannot operate with third-party technology. This is a significant barrier to innovation. Although not a significant technical challenge, it requires a greater commitment from all suppliers to integrating third-party solutions quickly. **NHSX** should work with software providers to ensure that this happens.

**5. NHSX should work with regulators and other national bodies to improve the data and related infrastructure required for clinical validation.** It will require larger datasets that have the right metadata and represent the population that the tool will be used for. It is important for effective validation and public trust that the validation data is not available to developers for training algorithms.

g. Excluding screening programmes where this work is performed by Public Health England and the National Screening Committee.

**6. CQC needs access to technical validation to underpin its regulation.** CQC will need to register and regulate organisations supplying autonomous machine learning systems as services. However, CQC does not have the technical knowledge to assess whether machine learning devices are providing safe and effective care. **NHSX** should explore whether technical validation can be provided by another part of the system with greater expertise. CQC could then refer to this in its regulation.

## 6. Conclusion

A level of risk is emerging that requires some tangible work in the short term, so that we can protect patients. This work is technically outside the capacity of CQC. However, there are resources, technical people and groups working in and with the UK notified bodies, the AI lab, HDR UK, the five centres for excellence in AI Radiology and Pathology, NICE, NHS Digital and NHSX. There is close work planned with industry around real-world testing and launch, which would benefit from coordination with regulators. This would provide a useful testing ground for CQC and MHRA to help direct activities, enable us to close these information gaps, and develop our approaches to regulation. There may also be some other approaches, such as setting up an independent accreditation body for data-driven technologies, which may address a number of these gaps if established in the right way.

Addressing these gaps will take some time. The market and industry are developing at pace. It is therefore important that we can respond as regulators when there are issues with data-driven technologies in health and care, even before we have these recommendations in place. To do this effectively, it is suggested that NHSX consider chairing a data-driven technology oversight committee that includes the MHRA, CQC, NHS Digital, AAC and, when required, relevant standards-setting bodies such as royal colleges and NICE, which can respond to emerging risks and issues quickly, and coordinate short-term technical work between organisations.

## 7. How we carried out this regulatory sandbox

### 7.1 Partners and activity

Seven technology suppliers and their NHS partners who were delivering machine learning applications in diagnostic pathways won a competitive application process to enter the sandbox, by demonstrating that they were at an advanced stage with real world implementation.

Members of CQC's staff formed a team from across different functions, including our National Professional Advisor for surgery.

To oversee the work, we also formed a governance committee with input from deputy chief inspectors from each CQC directorate, chaired by a member of CQC's Board.

The technology suppliers and care services using digital triage were:

- Dartford and Gravesham NHS Foundation Trust and Behold.AI
- East Cheshire NHS Foundation Trust and Oxipit
- EMRAD consortium of NHS Trusts and Kheiron Medical
- University Hospital Southampton NHS Foundation Trust and Aidence
- IBM
- Healthy.IO
- Sensyne Health and Royal Berkshire NHS Foundation Trust.

NICE, NHSX and MHRA were included as government partners in this sandbox round, and we have been working to scope new guidance for NHS providers on AI systems with the Information Commissioner's Office.

We also benefitted from the contribution of Jennifer Pearl, a CQC Expert by Experience with knowledge of using services, the Royal College of Radiologists, and the BSI's standards division.

We are grateful for advice and contributions from Google Health, UKRI, HDRUK, ICO, Cancer Research UK, ResApp, Qure.AI, Feedback Medical, NHS Horizons, UKRI Centre for Doctoral Training in Artificial Intelligence for Healthcare (Imperial College London), and the 2020 cohort of the LSE MPA Capstone project.

We worked together as a team to set out what good looks like and to understand the issues involved in services that use using machine learning.

After developing some questions for assessing the quality of care in services, we carried out a site visit for each participant<sup>h</sup> to refine our understanding of the issues and what is important. We then reviewed our regulatory approach, developed draft guidance, and held detailed discussions with our government partners.

## 7.2 Evaluating and learning from these pilots

We are running three pilot regulatory sandboxes in 2019/20, supported by the Regulators' Pioneer Fund through the Department of Business, Energy and Industrial Strategy. The purpose of the pilots is two-fold:

- to learn about what good quality means for emerging service types and how to regulate that
- to determine whether regulatory sandboxing should be part of how CQC works.

Once we have completed all three pilots, we will evaluate them and publish the learning and how we can use this to improve how CQC works in the future.

## 8. What good looks like – detailed statement

The first output from our regulatory sandbox process is a common understanding of what good looks like. By this, we mean the things that should be present to help deliver high-quality care when using machine learning applications in clinical diagnostics. Developing this shared view of quality – with people who use services, providers, technology suppliers and system partners – has been the basis of our work in the sandbox.

The following reflects the consensus position reached through the sandbox process across the five key questions that underpin CQC’s definition of quality: are services safe, effective, caring, responsive and well-led? This was adapted and developed after direct assessment of the services using these applications.

The responsibilities of machine learning suppliers, and providers that use the applications, may shift depending on the detail of the contractual arrangements in each case.

### **A note on defining a publicly acceptable performance benchmark**

When building consensus with providers, technology suppliers and people who use services, we established that the tool should be “at least as good as the relevant clinician it is seeking to replace”.

However, when thinking how to set this benchmark in practice, it was challenging. For example, for detecting lung nodules on X-rays, should it be as accurate as the average clinician in each hospital where it is deployed, or should some national average be used? Do we want to consider a general radiologist or a specialist? Further, are we only interested in the sensitivity and specificity of detection and diagnosis of the clinical feature, or are we also interested in higher order clinical skills such as evaluation and triangulation, and how well clinical questions are answered?

We suggest that suppliers, national bodies and providers should take a pragmatic approach, use available benchmarking data, and consult people who use services where possible. National bodies should work to develop clear benchmarks for different categories of device.

**CQC key question**

**Machine learning (ML) technology suppliers**

**Healthcare providers using ML solutions**

**Safe**

Good **technology suppliers** strive to deliver a safe product for all service users who they support.

They validate the performance of their algorithms where there is significant clinical risk, under the supervision of the clinical lead, and do so for each trust they deploy to, where necessary. They continue to monitor performance.

They test their algorithms for bias across different groups and are transparent about their performance in this regard. *This will require national bodies to facilitate access to enough metadata of sufficient quality.*

They work openly and flexibly with their customers to support, develop, and monitor robust clinical pathways of care.

They only release autonomous clinical AI systems when they perform at least as well as the relevant performance benchmark (for example, as well as a competent radiologist if seeking to automate lung nodule assessments).

They consistently implement a process for where clinician and ML readings disagree, so that the learning of the ML system is embedded.

Good **providers of clinical services** that are procuring ML systems have a managed approach to clinical deployment and use of ML applications.

They develop and consistently implement a process for where clinician and ML readings disagree, so that the clinical team learns from this, and clinical risks with the software are understood.

They work to map, plan and execute a clinical pathway that is operationally robust, and that involves clear and credible contingency plans.

They take steps to ensure the clinical leadership in the department has the clinical time and data that they need to validate the algorithms before deployment.

They conduct thorough clinical reviews in each clinical pathway where ML is deployed, to establish the strengths and limitations of the software and the desired level of autonomy from clinical reading.

Where primary reading occurs, or ML systems are reporting autonomously from clinical reading, services introduce this with a period of shadow reporting that supports learning by flagging where ML and clinician readings disagree. There should be an agreed proportion (e.g. 8-12%) that is continually double reported by clinician and ML system after deployment, and providers will satisfy themselves that the ML system performs no worse than existing clinical practice where they allow the system to run autonomously.

**Royal College of Radiologists teleradiology standards we think are relevant:**

- 1: There should be clear and transparent systems in place for rapid, secure transfer and review of images and where necessary storage of patient data.
- 2: Reporting must be the same standard independent of where and by whom the data is reported.
- 3: The same person should interpret the examination and issue the report to the referring clinician and should be clearly identified, with the results communicated and integrated into the base hospital's radiology information system (RIS), picture archiving and communications system (PACS) and electronic patient record (EPR) in a timely manner.



## Effective

Good **technology suppliers** strive to deliver a clear and accurate diagnostic service.

They ensure that their systems, people, and processes are implemented effectively for each customer site, based on their local requirements, and in the case of population screening programmes, in compliance with national standards required of services.

They work with customers to ensure that the pathways in which their solutions are deployed are effective for the whole system. For example, they do not lead to unwarranted increases in demand or an unacceptable drop in the rate of patients identified with a significant illness.

They have trained their algorithms honestly, and are transparent about the datasets they have used for training.

They will develop against an intended use that leads to a valuable solution for people who use services or clinical teams working in a clinical pathway.

They remain within their intended use as a medical device.

They continue to optimise their solution with training data, and this activity is expected to happen in parallel and separately to live clinical services. Updates will be released incrementally. Each update will be validated before being deployed.

They have trained and separately validated their systems sufficiently (with enough data that is separate between training and validation sets). Ideally the validation is conducted by an independent third party.

Good **providers of clinical services** that deploy ML solutions test and monitor the effectiveness of those solutions, ensuring that tech suppliers have done what they need to, and that they are used appropriately as part of clinical pathways.

They will have agreed to use the solution because of an identified need for people who use services, or a problem in their service. The solution will be proportionate to the problem that they are facing. They will have thought through the impact on people who use services and staff, and involved them in the redesign of the relevant diagnostic pathway.

They will audit the effectiveness of tools in their services, and support technology suppliers to validate and improve their systems.

They participate in national audit and learning around the use of these systems.

**Providers of screening services** will deploy machine learning in compliance with SQAS and PHE guidance and standards.

## Caring

*When using CQC's assessment framework, we will not inspect or rate the caring key question for machine learning technology suppliers unless they have direct contact with people who use services. However, as a group, we agreed the following as good practice:*

Good **providers of clinical services** deploy ML solutions in a service that cares for people and meets their emotional needs.

They explain to patients the next steps in terms of the care or diagnostic service, and how the ML solution works as part of their journey.

|                   |  |   |
|-------------------|--|---|
|                   | <p>Good <b>technology suppliers</b> support clinical services to care for patients by:</p> <ul style="list-style-type: none"> <li>• setting out their reporting in a clear and transparent way that builds trust</li> <li>• supporting customer-providers with materials for patients to help explain the system, how it works, and what the results mean.</li> </ul>  | <p>Their patients receive results in a clear context and have an explanation of what they mean.</p> <p>They explain to patients how their data is used, including secondary uses, the contribution their data makes to improving care for others, and provide a facility to opt out of sharing their data for secondary uses.</p> <p>They actively manage and consider people’s anxieties around diagnostic results, both in how they establish pathways of care, and how each patient receives their results.</p>  |
| <b>Responsive</b> | <p><b>Good technology suppliers</b> support clinical services to meet people’s needs. They take feedback seriously and set out their complaints process clearly online, contributing to NHS-wide learning from complaints.</p> <p>They provide results in a timely way and in clinical workflows that have been designed with their customers (hospitals) to reduce the risk of delay for people.</p> <p>They actively monitor for, and reduce, bias in their systems (unfairness). This will require national bodies to facilitate access to sufficient metadata of sufficient quality.</p> | <p><b>Good providers of clinical services</b> deploy ML solutions in a service that is fair, fast, and responds to people’s concerns.</p> <p>They use the ML solution to improve how quickly they can provide diagnostic reports to people who need them, especially to those who need them the most.</p> <p>They help their suppliers to actively monitor for and address bias (unfairness).</p> <p>They take an interest in, learn from, and share complaints between the service and tech supplier.</p> <p>They work to ensure robust continuity of care to help meet people’s needs in a way that is convenient.</p> <p>They provide information in an accessible format for people who use services.</p> <p>They respect people’s informed choice.</p> <p>They continue to take into account the needs and preferences of individual people, including those with protected characteristics under the Equality Act or in particularly challenging circumstances.</p> |
| <b>Well-led</b>   | <p><b>Good technology suppliers</b> are well-led when they:</p> <ul style="list-style-type: none"> <li>• engage the doctors/clinicians in design and configuration choices</li> <li>• develop systems with clear accountabilities at each stage, which are auditable</li> <li>• maintain a culture that promotes the interests of people using services,</li> </ul>  | <p><b>Clinical services</b> buying or using ML in their pathways are well-led when <b>they embrace innovation in a safe and well-informed way.</b></p> <p>They follow NHS Digital’s guidance on managing clinical risk from digital technologies (DCB0160).</p> <p>They ensure that suppliers follow the equivalent (DCB0129), are a licensed medical device, and that their CSOs communicate openly and allocate clear</p>   |

and highlights and learns from safety issues

- act transparently and openly with customers, clinicians and people who use services
- operate to best practice technical standards, including having robust back-up and governance processes around data and cyber security
- manage and monitor the performance and outcomes of the solutions they are developing
- have a clear vision for what the technology will bring to people who use services
- manage the funding cycle and related pressures so that quality is protected.

accountabilities to help identify, grade, and mitigate clinical risk.

They work with the PACS and RIS suppliers and specify contracts to ensure that machine learning solutions, outside of a pilot or testing settings, are integrated into the relevant clinical systems.

They procure systems carefully, and in doing so verify their regulatory approvals, clinical claims, and technical vulnerabilities, as well as working with the supplier to specify clearly how the solutions should be used and what additional conditions and dimensions of performance require contractual monitoring and management.

They do not block AI or similar projects by failing to understand the basic requirements and value proposition.

They develop close and productive working relationships between clinicians, managers, and developers. There is a clear division of accountability between management and frontline staff, and between the clinical service and the tech supplier.

They understand tech suppliers' quality management systems and have their own approach to improving quality.

This work was made possible by a grant from the £10m Regulators' Pioneer Fund launched by The Department for Business, Energy and Industrial Strategy (BEIS) and administered by Innovate UK.

The fund enables UK regulators to develop innovation-enabling approaches to emerging technologies and unlock the long-term economic opportunities identified in the government's modern Industrial Strategy.



Department for Business, Energy & Industrial Strategy

## How to contact us

Call us on: 03000 616161

Email us at: [enquiries@ccq.org.uk](mailto:enquiries@ccq.org.uk)

Look at our website: [www.cqc.org.uk](http://www.cqc.org.uk)

Write to us at: Care Quality Commission  
Citygate  
Gallowgate  
Newcastle upon Tyne  
NE1 4PA



Follow us on Twitter: @CareQualityComm

CQC-455-032020