Deloitte.

Improving Customer Outcome Testing

A practical guide for Boards

Contents

Overview: Consumer Duty intensifies the focus on customer outcomes

Under the FCA's (Financial Conduct Authority) Consumer Duty (the Duty), there are clear rules and regulations requiring firms to deliver good customer outcomes. In the early days post-implementation of July 2023 we have seen a more proactive and intrusive approach from the FCA with regards to firm conduct. Whilst expecting firms to measure and test customer outcomes is not a new concept, the regulator now has greater powers to act where it finds customers have not received "good" outcomes. The FCA has set out its intended approach on the Consumer Duty *'Our supervisory and enforcement approach will be proportionate to the harm – or risk of harm – to consumers, with a sharp focus on outcomes…We will prioritise the most serious breaches and act swiftly and assertively where we find evidence of harm or risk of harm to consumers.'* Sheldon Mills, Executive Director, FCA, May 2023.

The FCA have said that *'A key part of the Duty is that firms assess, test, understand and evidence the outcomes their customers are receiving. Without this, it will be impossible for firms to know that they are meeting the requirements set out in the Consumer Duty'*. The activity undertaken and type of information used will vary depending on factors like; size of the firm; target market; types of products and services; and distribution channels. However, this is typically likely to include the use of key data/metrics, outcome testing conducted in 1st and 2nd lines of defence and other assurance activity. The new Consumer Duty requirement for Boards to report annually (from July 2024) on whether they are delivering good outcomes highlights the importance of having a robust approach to demonstrate the outcomes being achieved, which should include an outcome testing framework as one of the key tools to support the ongoing monitoring and in turn, the Board report.

Outcome testing will be an integral pillar in the toolkit to test the outcomes achieved for customers through a customer journey and product lens, identifying any root causes of harm and prompting proactive action to fix, mitigate and remediate customer harms. This in turn will support that they have taken reasonable steps to deliver good customer outcomes. We often see an inconsistent or narrow approach to the testing of customer outcomes, without sufficient coverage of all key customer harm scenarios and with varied and/or limited sample sizes. This inconsistent approach to the testing methodologies and/or volume reduces the effectiveness of the testing, the ability to identify harm and the level of assurance provided.

The absence of an effective framework creates a risk that issues remain undetected and foreseeable harm will crystalise without providing firms sufficient time to act to rectify issues before they become systemic. Weaknesses in a firm's approach to outcome testing can result in poorer outcomes for customers, higher remediation costs and increased scrutiny from supervisors. We have observed that where potential systemic issues are identified by the FCA, this can often result in regulatory intervention which may include a Skilled Person Review under section 166 of the Financial Services and Markets Act 2000.

This point of view, which is based on our experience of designing and implementing frameworks alongside performing outcome testing on behalf of both regulators, clients and as part of our outcome testing outsourced service, sets out practical guidance to firms on reviewing and enhancing their current approach. Walking through the main stages of the outcome testing process, we set out key considerations and provide practical examples of the questions and challenges Board members and those responsible for customer outcomes can raise to assure themselves that the overall framework and approach is robust and fit for the purpose.

What is outcome testing, and what is it not?

Customer outcomes can be assessed using a range of data/metrics, customer testing and assurance activity. This point of view focuses specifically on outcome testing, which refers to a holistic review of a customer's journey to determine whether, based on their individual circumstances, they received a good outcome. In assessing whether the outcome was good for the customer in question, outcome testing goes beyond traditional Quality Control (QC) methods (such as point-in-time testing and training and competence schemes) which typically test whether internal policies and processes only have been followed and tend to be more employee competency focused than outcome focused.

Outcome testing can be undertaken at individual stages of the customer journey in specific functions (for example, complaints, customers in financial difficulty, fraud claims), as part of product or service reviews or through targeted thematic testing (for example, vulnerable customers and bereavement journeys). Depending on the scope and nature of the review being performed, outcome testing may consider factors including whether:

- The standard of communications and contact between the customer and the firm was appropriate, but also likely to be understood by the customer;
- The firm obtained and recorded sufficient information to understand the customer's circumstances (including identifying indicators of vulnerability);
- Products or solutions (e.g., forbearance) provided to the customer were appropriate and performed as expected;
- The firm acted appropriately at critical points and there was no undue friction in the customer journey (for example, during the advice or claims process); and
- The firm took reasonable steps to avoid foreseeable harm (for example, acted with sufficient pace when notified of third-party fraud to avoid further losses or under distress and inconvenience).

In practice, firms may apply, as the initial basis of their assessment, high-level regulatory expectations regarding conduct of business and treatment of customers and/or internal standards. However, it is vital that outcome testing is more than just a detailed gap analysis against the FCA rulebook or a check for adherence with internal policies and processes. Firms may apply their processes correctly and still deliver a poor outcome. Conversely, firms may decide that they need to adjust their standard processes or provide scope for exceptions to them in clearly defined circumstances to deliver a good outcome for the customer.

Another common misconception is that outcome testing is an assessment of customer satisfaction. Customer satisfaction may be indicative of a good outcome, however firms need to be mindful that customers can be satisfied with a poor outcome and dissatisfied with a good one or that customers may be unaware what outcome they ought to have received. In assessing the outcomes delivered for customers against the requirements of PRIN 2A.6 Consumer Duty: retail customer outcome on consumer support, firms will need to calibrate and determine their risk appetite in respect of when a suboptimal experience translates into a poor outcome. There can be a fine line between the conclusion that the firm could have improved their service versus the firm has caused harm through their process.

Do	Don't
• Consider whether the customer journey led to a good overall outcome based on the facts of the interaction(s) • Check that activity took into account the customer's individual circumstances and these were appropriately established • Focus on whether the product, activity or response was right for the customer as well as whether a process has been followed	• Focus solely on adherence with policies/process as these may not always deliver a good outcome • Limit the exercise to a gap analysis against detailed FCA rules • Assume that customer satisfaction is evidence of a good outcome

How does outcome testing fit with other controls?

We typically see outcome testing as one of the primary tools used by both firms and the FCA as a means of assessing, monitoring and importantly evidencing that good outcomes are being delivered for customers. Outcome testing should be incorporated into the wider suite of tools, data and controls to assess whether firms are delivering good outcomes for customers and used to inform improvements to products, services and the customer journey.

Firms' outcome testing frameworks should be informed by a wide range of internal and external inputs. There will be some customer interactions or journeys that are not suited to outcome testing and therefore other information/data or Management Information (MI) will be required to support firms understanding if these are likely to be causing harm or whether operational controls are working as intended. Section 11.33 of the Consumer Duty Final Guidance (FG22/5) provides illustrative information firms may want to collect. This list includes but is not limited to; business persistence, distribution of products/pricing and fees and charges, behavioural insights, staff training and competence records, complaints (number of and root causes) and drawing on external sources of data about consumer outcomes. As a starting point, Firms should compare this information against their own risk appetite and approach, considering whether further information or data points are required.

Whilst regular outcome testing of key customer interactions and processes are typically performed by the first line, the second and third lines may also use outcome testing to independently assess the effectiveness of controls and/or to identify thematic issues. There is a need for defined and robust governance to oversee the outcome testing process and results to ensure that Board Members and Executives have adequate visibility of customer outcomes, can set and monitor action taken to address issues and provide direction back to the business based on trend analysis or concerns.

The different methods of outcome testing

We set out below the different methods of outcome testing we would expect to form part of an outcome testing framework:

End-to-end outcome testing assesses the entire customer journey from the beginning of the initial interaction to the end outcome associated with that interaction, covering all calls and written communications. End-to-end outcome testing can be time and resource intensive. It can also be challenging to perform where aspects of the client relationship are handled by third parties, for example, intermediated sales. However, end-to-end outcome testing is a key tool to identify weakness in the overall customer journey that may have been missed by other controls and point-in-time testing. We would deploy this where journeys have defined start and end points. For example, an insurance or fraud claim.

Fixed-period outcome testing has a specific time/interaction period selected (for example, three months) and reviewers assess all interactions, records and documentation relating to the outcome in that time period. We would deploy this for customer journeys that may not have reached an end point during the assessment period. For example, collections and recoveries.

Point-in-time (PIT) outcome testing focuses on reviewing an individual interaction or aspect of the customer journey (for example, the initial sale) to assess the critical stage where a poor outcome could occur. Firms need to be careful that point-in-time testing does not become just a test of policy adherence rather than meeting the needs of the customer (which will not always be the same as customer wants) during the journey.

Thematic or "deep dive" reviews are helpful in assessing cross-cutting issues, root causes or areas of regulatory focus, for example, the identification and treatment of vulnerable customers. Thematic reviews may be either point-in-time or end-to-end reviews. However, a combination of both approaches is, in our experience, the best way of identifying thematic issues. Thematic testing is particularly useful where firms have different outcome testing methodologies in different business units making it more challenging to identify common areas of potential harm. Or to test hypothesis of where harm might be occurring but is not being identified through the business-as-usual testing.

The different methods of outcome testing

We set out below how different types of outcome testing may be performed across typical customer journeys for, regulated mortgages (from origination to exit), insurance, fraud alert management and fraud case management (investigation, and resolution).

Figure 1: Regulated mortgages from origination to exit example:

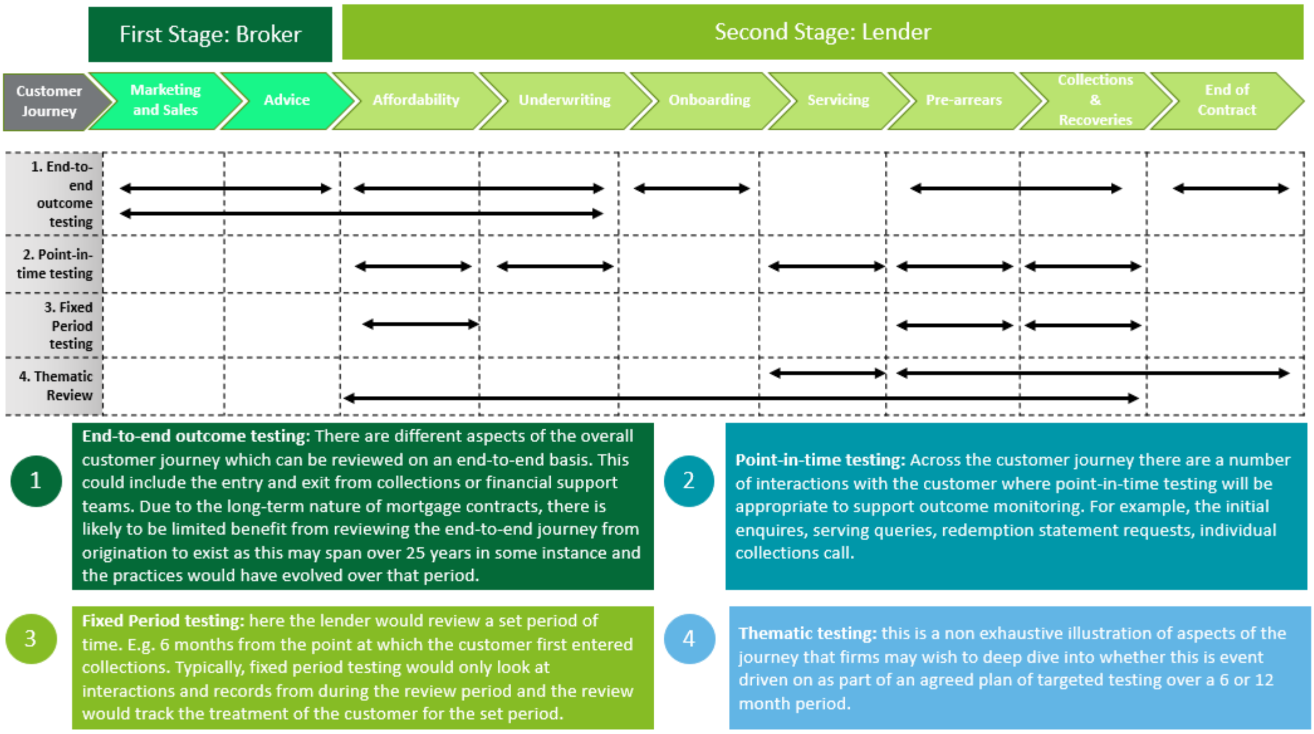


Figure 2: Insurance example:

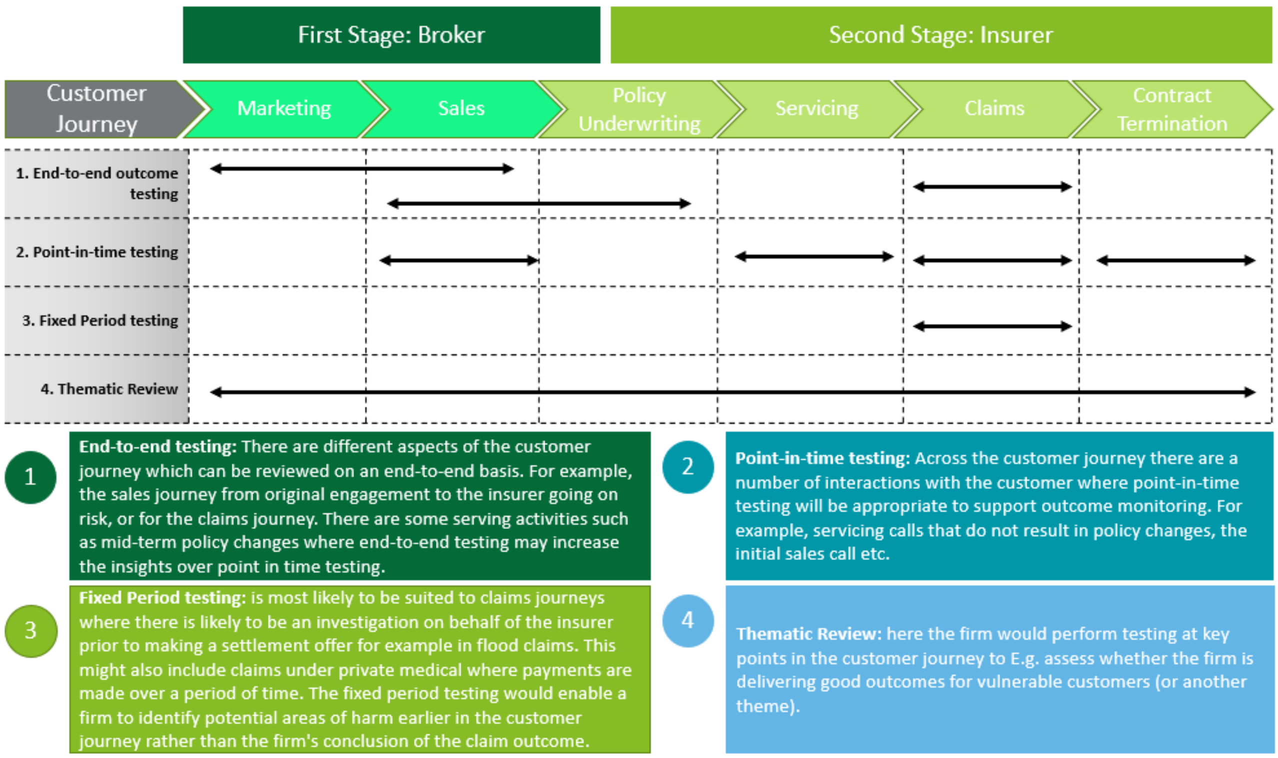
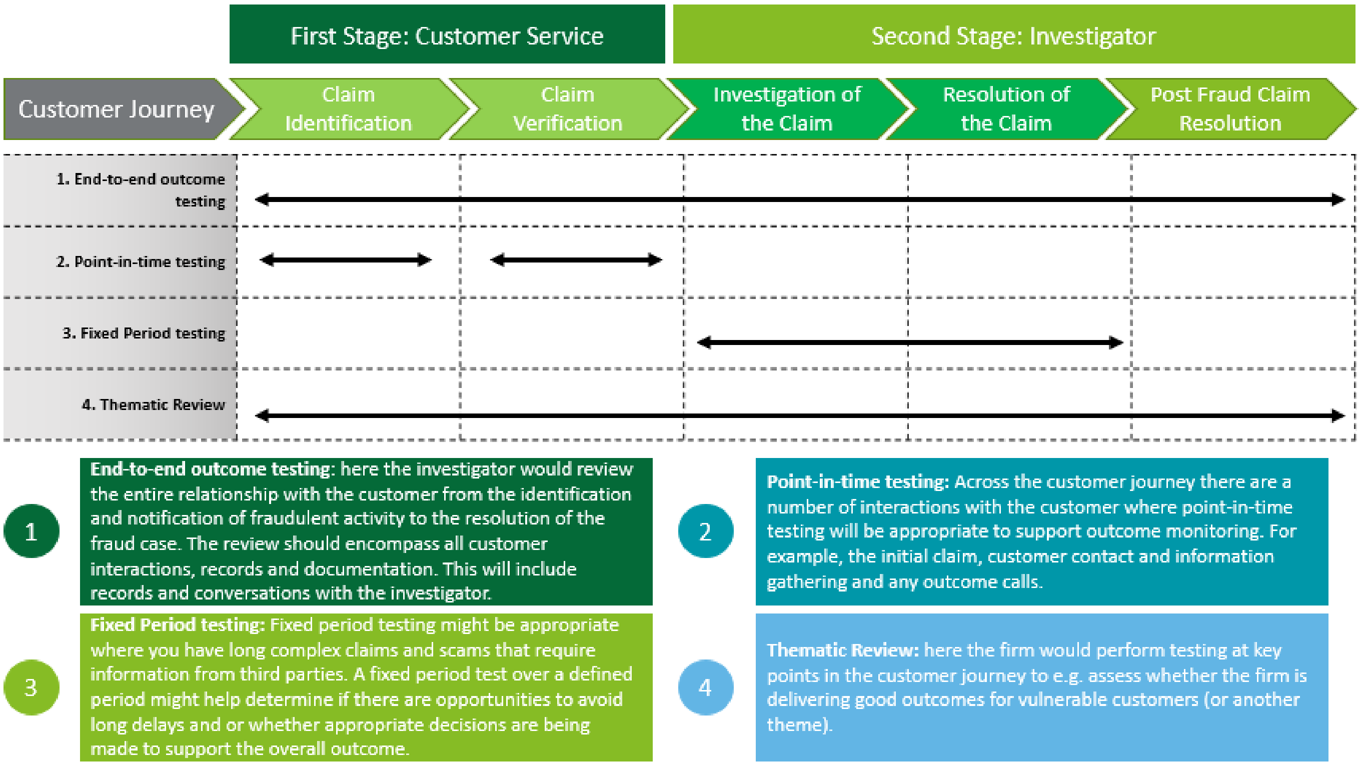


Figure 3: Fraud investigation and resolution example:



Monitoring customer outcomes under the Duty

We have set out in this section, the core components to support firms in delivering outcome testing. This includes selecting the sample, the testing methodology and determining the outcome for a customer, highlighting key considerations for firms. We also provide examples of the questions and challenges that Board members and Senior Executives can raise to gain assurance that the outcome testing process is robust, focused on understanding customer outcomes and designed with key supervisory priorities in mind.

Sample selection

Many firms grapple with determining what is the appropriate volume of outcome testing to undertake. In our experience, firms' primary consideration (when deciding how much testing to undertake), is often the availability of resource and therefore capacity within the outcome testing function (whether this is centralised or devolved to individual business units). This can potentially constrain the coverage of the testing (in one business area or across a firm), the choice of methods used (e.g., end-to-end versus point-in-time) and therefore the assurance provided. Rather than resource-driven, the overall outcome testing framework, volumes and approach should be determined by a firm's risk appetite, considering the wider data and MI available (which may help provide broader assurance) and be proportionate to the size and complexity of the business, including consideration of the customer journeys or products that have the highest potential for harm. This is crucial to ensure that Board Members and Senior Executives are able to place reliance on the results in demonstrating good outcomes.

CASE STUDY – Insufficient sample size

A firm undertakes fixed period testing on a monthly basis on its complaint handling function. The sample size is 10 per month, and the closed complaint volumes are around 800 per month. The sample size means that if one case is concluded to be a poor outcome for the customer, this is outside of the firm's risk appetite of 95% good outcomes across their complaint handling.

Where this is an ongoing trend there is a risk that senior management do not take reasonable steps to address any of the root causes, as the low volumes of poor outcomes means it is difficult to determine trends or if these represent systemic issues that need addressing.

In determining the type of testing that is appropriate to be deployed firms should consider this in the context of the inherent harms that are likely to exist for the product/service and/or interaction (recognising that fixed period or end-to-end might not be appropriate for all products/services and/or interactions) and the control environment.

Clearly defining the scope of assurance activity and testing approach should enable firms to ensure testing objectives are clearly articulated, understood and therefore achieved. Without a well-defined scope, testing activities can become unfocused and inefficient, and are less likely to provide valuable insights that can enhance and improve the customer journey or identify foreseeable harm.

What is tested and how it is tested will differ between firms and customer journeys depending on a number of factors including the risk profile but also the underlying processes. Specifically, the extent to which processes which rely on human intervention which have a greater execution risk compared to automated processes or those with hard controls.

Regardless of the testing methodology deployed, it is important that firms are clear what harms or risk the outcome testing is seeking to provide assurance over, otherwise there is a risk of senior managers taking false comfort or issues escalating that are more administrative in nature and have low risk of harm. Firms also need to ensure that there are clear requirements to selecting a sample of cases for review and that testing undertaken provides intended results. If parts of the customer journey are grouped together in one outcome test, the results may not accurately identify the root cause of the issue.

For example, in respect of a mortgage sales journey if the assessment includes the initial enquiry through to completion of the mortgage, is it sufficiently clear that the assessment includes both the suitability of the advice and the execution of the post advice aspects of the journey? If the outcome testing covers from the point of the initial client conversation to the product being loaded onto the system, if the wrong product is recorded onto the system, are senior management provided sufficient visibility to support them understand poor outcomes. This does not necessarily mean that the advice provided to the customer was unsuitable, but there are potentially operational issues. Similarly, if the complaints outcome testing only commences once the complaint is passed to a centralised team but does not include the initial conversations with the customer, is it sufficiently clear that the assurance is limited to the central complaints function's ability to execute based on an internal team's instruction rather than whether the investigation covers all points raised by the customer?

In developing and agreeing the programme of outcome testing activity, firms will want to pay particular attention to likely areas of supervisory focus. The Duty has highlighted the four outcome areas where the regulator is expecting increased attention. Outcome testing can be used as one way to monitor increased attention, however, standalone, it does not provide sufficient evidence that firms are satisfying these outcomes.

It is also essential that the outcome testing programme covers a sufficiently wide spread of customer cohorts, products and/or product features, with particular emphasis on those that have the greatest potential to avoid foreseeable harm and the most adverse impact overall. Firms may find it helpful to use a risk assessment framework to identify higher risk products/features for inclusion. To ensure that their testing is robust and does not miss issues during different points of the customer journey, firms should ensure that they use a combination of different methods (see page 5).

Having selected the focus areas and the method (or combination of methods) for a particular review, firms need to identify an appropriate sample of customers. Firms need to be able to balance a number of considerations:

- The resource required to review a sample sufficiently large or statistically significant to enable the results to be extrapolated across the book of business;
- A sample that is likely to indicate potential systemic issues and provide validation to the trends and insights from the conduct risk MI.

The Duty brings an additional complexity to the sample selection process, as firms are required to monitor the outcomes being achieved for different customer cohorts, including vulnerable customers and customers with protected characteristics for example. A lack of good quality customer data and analytics can frustrate firms' attempts to select a representative sample, requiring manual, case-by-case checking. However, data or cluster analysis can assist firms in identifying appropriate cohorts of customers for thematic testing. The FCA expects that firms will consider outliers and undertake deeper analysis to determine that harm has not crystallised.

Table 1: Sample selection considerations

	Key considerations	Key Questions for the Board and Executives
Sample selection *Choosing a proportionate, representative sample*	• The proportionality of the outcome testing framework to the size and complexity of the business.	• As a percentage of our customer base, how much outcome testing are we performing annually? How does this compare to figures for other forms of Quality Assurance (QA) and first line testing? • Are we using an appropriate combination of outcome testing methods so that the full end-to-end customer journey is covered? • Is our outcome testing programme designed to assess current supervisory priorities? If so, which ones? • Is there sufficient coverage of products/product features, particularly those deemed high risk?
	• Resourcing of the outcome testing function is driven by factors including the size of the outcome testing programme, the methods of testing to be used and the areas/products in focus.	• Are the teams performing outcome testing adequately resourced to execute a proportionate outcome testing programme?
	• The representation of the customer base across different demographics and geographies including customers who are vulnerable.	• How have we assured ourselves that we have included customers experiencing a sufficiently wide variety of actual or potential vulnerabilities e.g., health issues, life events, financial hardship etc.?

File review

File review is typically the most time and resource intensive period in outcome testing as, to be performed accurately, all communications with the customer (whether by letter, phone, e-mail, webchat etc.) and electronic records should be reviewed in full over the defined timeline of the review (end-to-end, fixed period etc.). Restricting testing to a small number of customer interactions means a review can provide only limited assurance about the quality of customer treatment. Reviewers need, therefore, to be given an adequate amount of time to conduct their review; and firms should ensure that targets or timeframes around completion do not result in reviewers feeling pressured to "skim" the information available. Data analytics can be a powerful tool in helping firms process high volumes of information, and flagging potential issues for reviewers to take a closer look at.

It is vital that the review is undertaken in chronological order. This helps prevent the reviewer's judgement being influenced by the outcome the customer eventually received (rather than the outcome they ought to have received) and ensures reviewers assess whether the firm's actions were appropriate at the time they were taken, based on the information the firm had (or should have had) about the customer. Reviewers should, however, note any changes in customers' circumstances throughout the period being reviewed and assess whether the firm identified these and responded appropriately.

Ahead of undertaking the file review, firms should ensure that reviewers understand specific areas of supervisory focus and that guidance and toolkits have been amended to take account of these. Guidance and toolkits should also be adjusted to take account of the specific stages of the customer journey.

Table 2: File review considerations

	Key considerations	Key Questions for Senior Managers
File review *An independent review of customer interactions, electronic files and firm actions*	• The coverage of interactions with the customer and electronic records during the time period selected.	• What is our approach to file review? Are we checking all customer interactions during the relevant period? • If not, how have we assured ourselves that we are not missing key information? • Could investment in data analytics help us perform more thorough outcome testing?
	• The re-drafting of any guidance and toolkit to take into account the latest regulatory expectations and areas of focus.	• Do our reviewers understand the supervisory priority placed on issues such as customer vulnerability and fair pricing? • Are they able to identify them appropriately?
	• The time and freedom given to reviewers to examine each case thoroughly and in the right order.	• Do we set targets for the completion of reviews? • If so, do these allow sufficient time for reviewers properly to assess the information available?

Assessment and decision

This can be the most challenging aspect of the outcome testing process as reviewers are required to use their judgement to decide what, based on the customer's individual circumstances, constitutes a good (or poor) outcome. In our experience, some reviewers lack the necessary knowledge, experience and confidence to make a judgement whilst others adopt an overly "tick-box" mindset. To avoid decisions being little more than an assessment of whether the firm adhered to policies, procedures and rules, firms should ensure reviewers have appropriate training and guidance as to what constitutes a good/poor outcome.

Inexperienced reviewers may also fail to appreciate the material effect that seemingly small process or human errors can have on the overall outcome for the customer or lack the confidence to challenge the prevailing orthodoxy that the firm generally delivers good outcomes. As a result, their decisions can paint an overly positive picture of the treatment of customers.

Therefore, definitions of good and poor with examples/case studies are key to helping colleagues make those judgements and be consistent with each other.

CASE STUDY – Failure to follow process

A customer was a victim of third-party fraud who was convinced to make a large deposit payment to a builder via faster payment. On receipt of the funds the builder did not start the work as promised and the customer was unable to make successful contact with them. Consequently, the customer contacted their bank for help. The bank's own investigation established that they had not taken sufficient steps to warn the customer that the payment to the third-party could be a scam and refunded the amount of money in full. However, they did not delete the builder's payment details which is in line with the bank's process to reduce the risk of the customer sending further payments to the third-party.

When the case was subject to outcome testing, the tester concluded that whilst the customer had received a full refund, as the third-parties payment details had not been deleted from the customer's account this was a poor outcome.

Firms need to establish their own risk appetite and consider the balance between potentially over reporting on poor outcomes but providing clear operational guidance to testers. In this particular scenario, there are three key points here that are worth further consideration in reaching a conclusion. Firstly, the payment was not made in error or to erroneous details and do not need to be re-entered. Secondly, the customer contacted the bank for help having been the victim of a scam and did not receive the service in exchange for the funds being sent. On balance therefore it may be reasonable to conclude that the customer is unlikely to send another payment to the builder and therefore the risk of harm is limited. Lastly, if during the review period another payment has not been sent then no further harm has been incurred.

In the case of romance scams, where the bank as played a key role in alerting the customer to the scam, then deleting the payment details of the third-party is likely to be a more important factor in preventing foreseeable harm. Therefore, emphasising the need to scenario-based examples to support testers reach conclusions based on the customer's individual circumstances rather than policy and procedures.

High level targets which set the firm's risk appetite regarding good/poor outcomes can be a useful means of tracking, at senior levels, improvements or deterioration in the treatment of customers. However, proper oversight should be put in place to ensure they are not read by outcome testers as targets or "quotas" for the number of files they need to assess as good/poor.

To assist with QA later on, the reviewer's summary should be factual and concise and set out a clear rationale for their overall decision. In addition to noting any process/rule breaches and errors, they should note files containing insufficient information to make a decision as these may indicate wider issues around recordkeeping.

The role of second and third line can also be critical here in ensuring that reviews also do not become complacent of known issues in the relevant business unit or desensitised to these as they are frequently observed.

Categorisation

Given outcome testing is usually undertaken in different business areas across firms, there will often be inconsistent approaches to the categorisation of outcomes, including definitions, which impairs firms' abilities to aggregate assurance and outcome data and consolidate the view across the entire customer journey, product or business unit. The aggregation of rating definitions will ensure that management and stakeholders, including recipients of reports and MI will have clear understanding of outcome scores and what that means for customers and the business.

When determining an approach to categorisation, firms should also consider the number of possible outcomes. Simplifying the approach to include two outcomes e.g., good and poor makes it clear whether the customer has received a good outcome or not. (Or for investment advice whether the advice was "suitable" or "unsuitable"). A third outcome is sometimes used where the customer received a good outcome but there are opportunities for the firm to improve or where something went wrong in the period being tested but it was rectified and resolved by the end of the review period. We have seen little benefit of introducing and fourth outcome such as "unclear". This is often used in situations where information cannot be obtained to determine whether the customer received a good or a poor outcome, however if information cannot be obtained, this is a problem within itself and should be categorised as a poor outcome.

CASE STUDY – Scope of testing and driving consistency in reporting

As part of the ongoing collections and recoveries six-month fixed period testing, a lender considered the activity three months prior to the review period. Overtime this practice evolved to include considering any actions taken by the bank or the customer in the following month(s) to determine whether the customer has received a good or poor outcome. This practice was not consistent across the testers.

The lender would also consider whether there was known remediation in flight to address any issues that were identified, and consider these to be mitigations for any poor outcomes.

The deviation from the six-month period means that the lender does not get an accurate and comparable view of the outcomes being delivered over time. The inconsistent timeline being applied amplifies this and does not support the lender identify if there are any systemic issues emerging or whether controls are timely and effective.

Where there is known remediation inflight until this is executed it is more transparent to report the outcomes as poor, however the narrative supporting the MI would provide Management with the insight that further activity was unlikely to be required following completion of the remediation. Where customer accounts can be readily identified for specific remediation activity firms can consider removing these from the sample (subject to their internal governance) so this does not unduly impact the reporting. It also allows testers to focus on providing assurance on the wider population rather than continuing to report on known issues.

As firms are embedding the Duty we have observed that there is a greater need to calibrate the outcomes being delivered given the paradigm shift in expectation. This is also complicated as many journeys subject to outcome testing will straddle both a pre and post Duty timeline. The key aspects of the Duty driving the change in the outcome conclusions are:

- **Acting in good faith:** Is there any evidence that information has been deliberately or inadvertently misled the customer which has resulted them in taking action which impacts their ability to pursue their financial objectives?
- **Consumer Understanding:** Prior to the Duty coming into force firms have focused on whether the right information was disclosed to the customer during the relevant points in the customer journey. Firms now need to consider the information needs of the target market and whether the communications are likely to be understood. This is likely to be complex due to consumer bias but also firms need to apply caution to reliance on adherence to PRIN2A.5.9R (2). PRIN2A.5.9R (2) sets out that retail customers should be asked whether they understand the information and if they have any further questions, particularly if the information is reasonably regarded as key information, such as where it prompts the customer to make a decision. Firms will need to think about how they support those undertaking testing to recognise that customers might not understand despite saying they do. Where there is evidence that the customer does not understand the information being presented firms will need to consider whether frontline staff provided sufficient sign posting to third party support or encouraged a break in the process for the customer to reflect/research before determining the next step.
- **Consumer Support:** As noted earlier in this paper, firms will need to assess where a suboptimal customer experience results in consumer harm. Typically, pre the Duty we observed two approaches being adopted to assessing the customer experience as a whole as well as the outcome. Some firms would consider a poor experience to be a poor customer outcome and others would only consider particularly egregious or journeys were there was undisputed distress caused as a result of how the firm achieved the end outcome for the customer. The reality of what is good and poor is more nuanced. Firms will need to navigate and calibrate the extent to which journeys which could be concluded in a simpler manner are adding unnecessary friction or frustrating the customer's ability to pursue their financial objectives.

Table 3: Assessment and outcome

	Key considerations	Key Questions for the Board and Executives
Assessment and outcome *Reaching a decision on whether the customer received a good outcome overall*	• Reviewers' knowledge and experience of products and supervisory expectations regarding the good treatment of customers.	• How experienced are our reviewers? • Do our reviewers have an appropriate level of industry and firm experience to understand our products and expectations regarding the treatment of customers?
	• The extent to which reviewers are encouraged to demonstrate independence of thought, critically assessing whether the outcome was good, rather than rationalising the outcome the customer actually received. • The documenting of a clear rationale for the reviewer's decision.	• Do we provide any training or guidance to reviewers to help them arrive at a decision regarding what constitutes a good (or poor) outcome? • Have we, as Senior Managers, taken any steps to build our reviewers' confidence to the required level? For example, by communicating the value we place on outcome testing and the early identification of issues.
	• Oversight and governance of targets or defined SLAs regarding the treatment of customers and response times.	• Do we set targets/risk appetite to monitor the treatment of customers? Is there proper oversight in place to ensure the outcome testing function does not think of these as targets for the number of files to be found good/poor?
	• Consistent categorisation across business areas and units to ensure common understanding of outcome categories. • Simplifying the number of categories used and considering removing the category 'unclear' or equivalent.	• Is there a common understanding across our firm about what outcome categories mean? And is this the same across each business unit? • Do we currently have an 'unclear' or equivalent category? Is it being used as intended and is it adding value?

Quality assurance and root cause analysis

To overcome the inherent subjectivity in outcome testing, and potential inconsistency between individual reviews, firms should perform adequate QA. The most common forms of QA are sample checking and manager/peer review. However, case clinics and roundtables are also a useful means of identifying key themes, ensuring appropriate challenge of decisions/results and giving team members a strong understanding of the outcome testing process.

It's essential that firms assess the root causes of any failings or errors to determine whether they were isolated incidents or whether they indicate widespread/systemic issues regarding the treatment of customers. When identifying potential root causes, firms should look beyond factors such as training, policies and procedures, systems and controls and consider the extent to which the firm's culture (and sub-cultures), business model and remuneration practices may drive poor outcomes for customers.

The results of outcome testing must be used to inform improvements to both the outcome testing process and to products, services and operations. We often identify weaknesses in the feedback process due to a lack of confidence and autonomy in the outcome testing function. In some cases, the outcome testing function expects the business to validate its findings, threatening the independence of its reviews. More commonly, outcome testing functions can experience resistance to their findings and are insufficiently empowered to require the business to act and to monitor action items to closure.

The culture and mindset within the firm can also act as a barrier to identifying and reporting issues regarding the treatment of customers. For example, firms may have a "good news" culture which discourages staff from raising issues. There may also be an a priori assumption that the firm generally provides customers with good treatment, resulting in resistance to, or disbelief of, adverse findings.

Table 4: Quality assurance and root cause analysis

	Key considerations	Key Questions for the Board and Executives
Quality Assurance and Root Cause Analysis (RCA) *Identifying root causes and making necessary changes to policies, processes and procedures*	• The range of QA methods used to ensure consistency between different reviews.	• What themes are emerging out of our RCA? • Are we identifying issues beyond policies and systems and controls that need addressing?
	• The extent to which results are fed back to the wider business and used to inform improvements to products and services.	• What actions are being taken as a result of QA and RCA? • Are these being tracked? • Does responsibility for closing these actions sit with the appropriate teams?
	• The independence and influence of the outcome testing function.	• Does the business understand the priority the Board and Senior Executives place on addressing the issues identified by outcome testing? • Does the outcome testing function feel comfortable and empowered to raise issues regarding the treatment of customers?

Management information

Senior leadership need timely and effective MI. It is not possible for a board or senior committee to discharge both the collective and individual responsibilities held by members without reliable data. As well as these internal senior expectations, the regulator expects firms to have in place appropriate MI to test, monitor and demonstrate – both to their supervisors and to themselves – that they are delivering good outcomes for customers. In the absence of appropriate MI, firms will struggle to identify issues leading to poor outcomes and a failure to take effective action.

It is important that the visibility of outcome testing MI is not restricted to the first line or to conduct risk forums/committees and is used to inform operational and strategic decisions regarding the treatment of customers at various levels within the firm.

When reporting MI to senior committees, firms will, understandably, want to ensure that attendees are not burdened with cumbersome MI packs and so will summarise or aggregate findings. In doing so, however, they should be mindful that key insights can be lost. For example, presenting an overall figure for outcomes achieved across the firm can obscure poor results in specific business areas. Accordingly, when reviewing MI on customer outcomes, supervisors will expect the Board and Senior Executives to challenge the results. In particular, the Board and Senior Executives should be sceptical of, and challenge, results indicating consistently, unexpected or unusually high numbers of good outcomes, particularly if any accompanying narrative indicates failings in specific areas or at key points in the customer journey.

Information on trends can help the Board and Senior Executives track improvements or deterioration in the treatment of customers, as well as giving them a basis on which to challenge results (for example, if the trend shows consistently high good outcomes with little variation). Page 15 and 16 set out some examples of the MI that can be presented at various levels within the firm.

Table 5: Management information

	Key considerations	Key Questions for the Board and Executives
Management Information *Monitoring and reporting outcomes*	• The frequency with which outcome testing MI is reported to committees other than conduct/customer committees (e.g., Executive Committee, Risk Committee, Board).	• What MI are we seeing on customer outcomes? • Does it give us sufficient visibility and granularity on the key aspects of the business?
	• The balance between information that allows the Board and Senior Executives to monitor overall trends in the treatment of customers and the flagging of specific risks, issues or failings.	• Is the MI we see consistently positive? Have we challenged the outcome testing function to explain why this is the case and whether they are adequately highlighting any areas of poor performance? • Have we preformed any deep dives on outcome testing MI?
	• The tracking of action items including individual/functions responsible for closure.	• Does the MI show the action being taken to address any issues? • Are we monitoring whether these are being closed in a timely fashion?

Figure 4: Considerations regarding the presentation of outcome testing MI to internal audiences at different levels within a firm

Board/Board Committees

- Result of outcome testing should form a core part of any conduct, customer or operational management information.
- This may include aggregated figures of volume of fair/unfair outcomes across product categories and customer journey but should also focus on qualitative narratives on key themes, issues in specific functions, and trends.
- High-rated or widespread issues should be suitably reported or visible at this level to avoid underlying risks being lost in aggregation.
- Sufficient time should be allocated on the agenda to discuss points. Meeting minutes should demonstrate debate and challenge on specific MI and read across to other areas of agenda where relevant.
- MI should enable Board/Board Committees to consider and challenge any strategic changes required in products and services conduct risk appetite, with clear actions and accountability noted.

Senior Management/Compliance

- Detailed weekly/operational level MI may be complemented with aggregated information to identify key risks or trends across functional areas.
- Clear tactical and remediation activity is reported and considered for decision by those tasked with responsibility to manage customers on a day-to-day basis.
- Key trends from RCA are highlighted and actions identified and agreed.
- The results of thematic outcome testing or specific assurance carried out by the business is reported, as well as risk, compliance and internal audit.

Team Leaders/Heads of Function

Team Leaders receive regular suite of MI in relation to areas of responsibility based on testing conducted. This may include:

- Number tests performed
- Products/product features/teams covered
- %fair/unfair outcomes across testing activity
- Key trends and remediation types, including umber customer files requiring remediation
- Activity outside of appetite that requires escalation

In addition, Team Leaders receive ad hoc MI from thematic reviews identifying issues in their area of responsibility.

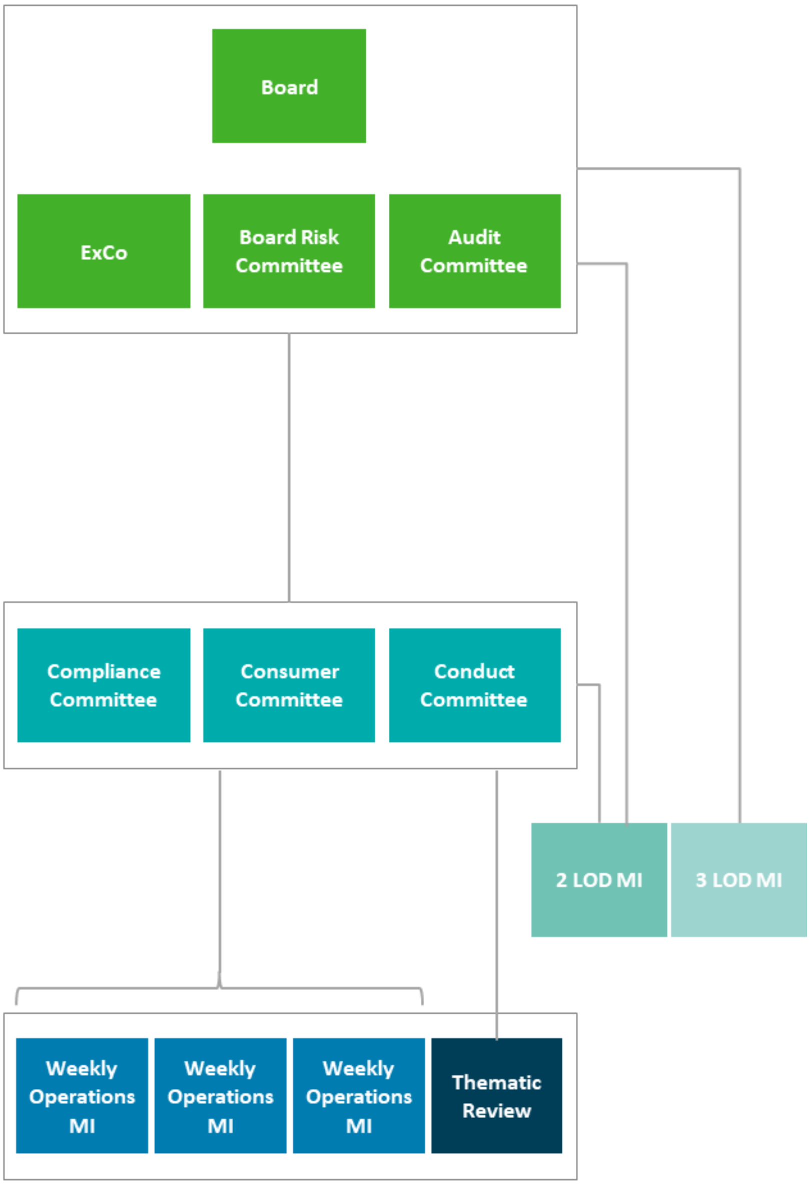


Figure 5: Fraud Workflow - outlines key roles and responsibilities across first- and second-line divisions to support the assessment of an organisation's fraud culture.

Data, Governance and Communication

Board/Executive Oversight

- Ensure MI and metrics are shared with executive colleagues to support testing requirements and enhance oversight capability of trends, risks and dependencies. Firmwide challenge and input should be encourages to promote a "customer first" mindset, with MI, KPIs and risk appetite statements aligned to this focus. Trends and metrics should be scrutinised in Committee meetings, as and where required.
- Executive level stakeholders should be aware of high risk and/or key themes identified in MI to support good customer outcomes, with a comprehensive view of key concerns, gaps and customer risk. This should be included in a formalized controls assessment to help define and balance a firm's fraud risk appetite with improved customer support.

2LoD

- Consolidated trainings on systems and processes should be provided in connection with outcome testing, including firmwide knowledge on Consumer Duty requirements. This could include fraud trends and leading practices on customer support, assistance, and resolution.
- A fraud strategy should be created within the 1LOD function, however they will be expected to provide dedicated oversight on the delivery of good outcomes. Appropriate KPIs and assurance reviews should support this, e.g., fraud reporting rate, chargebacks, false positives.
- High-risk and/or vulnerable customers should be identified through fraud case lookbacks and root cause analysis ("RCA")
- Fraud data and MI should be gathered and reviewed in conjunction with deficiencies found through the RCA exercise and risk appetite statements to identify key trends in functional operational areas.
- Testing and assurance procedures enforced by Compliance should be used to mitigate any deficient fraud controls affecting customer outcomes, considering upcoming regulatory change including the PSR liability shift.

1LoD

- MI and SLAs pertaining to outcome testing should be shared with relevant parties. This may include:
  1. Number of fraud cases resolved within a certain timeframe
  2. % of fair/unfair customer outcomes, customer complaints
- Sufficient investigation records are kept to allow post-fraud actions to be taken, e.g., firmwide communication or investigation of related alerts.
- An end-to-end case management tool should be employed within the fraud detection capability to enhance investigation tracking. This will ensure customers are kept appraised of any progress throughout the investigation cycle by a communication loop, meaning firms can manage and monitor customer outcomes through regular touchpoints during case investigations.
- Customer outreach should document how fraud events are decisioned to allow reimbursement and/or chargeback decisioning to be initiated. Any responses should be gathered and consolidated and fed into MI shared with 2LOD and executive colleagues to assess, test and challenge as required.

Operating model

We would expect as firms place greater focus on outcome testing that considerations will be given to how they deploy their testing teams and the operating model. The FCA do not seek to mandate how firms monitor the outcomes being achieved for customers. In particular, there is no defined expectation regarding the organisational structure deployed to undertake and oversee the testing undertaken by firms. There are a number of factors which drive the operating models across the different testing teams which include:

- Size and scale of the organisation;
- Complexity of products and services offered;
- Distribution and servicing channels;
- Broader governance structures;
- Historic structures; and
- Business maturity.

In federated business models where testing is decentralised firms need to consider how they can drive consistency in how they test and report so that this can enable effective aggregation that is both meaningful but does surface issues in individual business units. This is particularly challenging where you have differences in sample volumes or start and end points for testing.

Where we see greater variation is in the testing that supports business unit T&C schemes, which tends to be PIT testing. Firms operate a range of practices, including testing by team managers/supervisors, in-function testing teams, and centralised teams (either as part of the E2E and FP testing teams or separate). Many firms have moved between the models to optimise the role of the team manager or to manage the potential risk of conflicts of interest.

The centralisation of the testing has a number of operational benefits including (but not limited to):

- The ability to drive appropriate levels of consistency and efficiency in the approach to testing and reporting of outcomes;
- The ability to flex resource across the testing teams to support risk driven or themed testing as appropriate; and
- The independence from the business enables testing to consider the outcome for the customer even where this may span a number of functions or business units in the journey being tested.

For testing linked to the Competence Policy it is important that there is an appropriate feedback mechanism to support coaches and line managers with delivering feedback to colleagues, whether there is something to celebrate, consider, or do under the existing categorisations.

In considering the operating model, firms will need to consider the best structure and approach to support them in providing oversight of the customer outcomes through testing undertaken by the business. There will be a range of factors which are likely to influence the optimal structure, this includes:

- Where the capacity and capability sit in the short term;
- The level of changes proposed/required to the existing testing structure, approach and methodologies. For example, does this need to move from the business units to a central function in order to undertake the testing effectively;
- Whether there is centralised capacity/capability to support the business units respond in the event of foreseeable harms emerging;
- If testing remains in the business units, what level of oversight and centralised support can be provided to support the business units in defining methodologies and testing volumes to support them meet the minimums standards in the approach document; also
- Whether this remains inhouse or is outsourced on an ongoing basis or to support with surge capacity requirements and/or thematic reviews.

Outcome testing: what can go wrong?

Carried out well, outcome testing represents a valuable opportunity to identify improvements in the treatment of customers. However, we have identified a number of common challenges and areas of weakness and failings that can seriously impair an outcome testing process. These include:

Subjectivity	Subjectivity is an inherent risk within outcome testing as the assessment depends on customers' individual circumstances. To address subjectivity within their outcome testing, firms should set clear standards on what is a good/poor outcome, record clearly the judgement reached and why and ensure they have robust QA processes in place.
Culture and Mindset	Firms can approach outcome testing with the wrong mindset, viewing it as a "tick-box" exercise rather than a genuine opportunity to identify issues in the treatment of customers, and to improve products and services. The culture within a firm can also lead to an assumption that the firm generally delivers good outcomes for customers, resulting in resistance to any adverse findings.
Resourcing	Outcome testing is time and resource intensive and firms are often unwilling to commit sufficient resource necessary to undertake a testing programme of sufficient depth and breadth. However, a proportionate testing programme can help with the early identification of issues, reducing the risk of costly remediation further down the line.
Firefighting	Used regularly as part of firms' conduct risk management toolkit, outcome testing can help identify emerging issues in the treatment of customers. However, it is often only deployed once a risk or issue has crystalized, and as a means of assessing the scale of foreseeable harm that has already occurred.
Seniority and Independence	Outcome testing is often viewed as an extension of QA rather than an independent control in its own right. In addition, more junior staff can lack the confidence and gravitas to challenge the business effectively. Firms should ensure that they attach sufficient influence and practical status to the outcome testing function and that reviewers are empowered to challenge the business.

Conclusion

Monitoring whether or not a firm is delivering good customer outcomes on an ongoing basis will be a critical part of demonstrating compliance with the Consumer Duty. Outcome testing in isolation will not be sufficient but will be an integral pillar in the toolkit for firms to both test the outcomes being achieved for customers across the product lifecycle. Firm's resources will be finite, therefore a well-defined framework and appropriate deployment of testing (end-to-end versus point-in-time) and resource will be key to maximise the assurance and insights being achieved.

Authors

Kat Andrews Director katandrews@deloitte.co.uk +44 20 7303 2311	**John Lonen** Director jolonen@deloitte.co.uk +44 2920 26 4615
Emilie Rawlins Senior Manager erawlins@deloitte.co.uk +44 20 7007 3604	**Emma Lamont** Manager elamont@deloitte.co.uk +44 20 7303 7161

Key Contacts

Dave Clements Partner daclements@deloitte.co.uk +44 117 984 1199	**Lyndsey Fallon** Partner lfallon@deloitte.co.uk +44 161 455 6084
Chris Jamieson Partner cjamieson@deloitte.co.uk +44 20 7303 7996	**Alastair McGeorge** Director amcgeorge@deloitte.co.uk +44 131 535 7822