Evaluation Task Force

Single Customer Account Evaluation Plan

Evaluation Task Force

August 2025

PUBLIC

HM Revenue & Customs

Evaluation Task Force

Contents

Introduction

In 2021, the Single Customer Account was introduced as a key pillar of HMRC's digitalisation and automation agenda. The programme was allocated £105m to create an end-to-end digital service that addresses customer pain points, increases voluntary compliance and improves customer satisfaction. As of May 2024, the SCA programme has completed the development of the personal account system, has onboarded several core Child Benefit services, and implemented functionality for viewing personalised information in Pay As You Earn (PAYE) and National Insurance tax affairs.

This evaluation plan defines, through a Theory of Change, the key inputs and activities in building the SCA programme. Against the programme outputs, we have identified 20 outcomes and four impacts which can be used to evaluate the success of SCA. The Theory of Change is accompanied by a database of evaluation metrics, including the individual units, calculations and expected changes related to each outcome. A visual logic model details the causal flow between short-, intermediate- and long-term outcomes and their related impacts.

Through a Process Evaluation Outline, we set out a plan to gather lessons learned from the delivery of the SCA programme, from its inputs and activities in the development of the outputs, and the delivery of the outputs to the end-user. In this section, we provide a breakdown for conducting the proposed data collection methods, including interviews, surveys, focus groups and programme documentation.

In the Impact Evaluation Outline, we discuss the opportunities and constraints for evaluation that have determined our selected evaluation methods. Where it is possible to randomise the rollout of the SCA service to users, we suggest using A/B testing to evaluate these metrics. Given availability of sufficient data, we suggest creating a synthetic control against which treatment effects can be compared. Finally, where data availability is not sufficient, we suggest identifying an instrumental variable or conducting propensity score matching.

Through a Value for Money Evaluation Outline, we make limited suggestions to build on the detailed cost-benefit analysis of the SCA business case. In line with the National Audit Office's framework, we suggest a holistic evaluation that accounts for the economy, efficiency, effectiveness and equity of the SCA's implementation. Given the broad scale of this programme, we also suggest several potential unintended costs and benefits that merit scrutiny in the final evaluation.

Finally, through an Implementation and Timescales section, we discuss the wider considerations to be made for the implementation of the full-scale evaluation. In this final section, we discuss the skills and capabilities required for the evaluation team, an indication of timelines, and estimates of the cost for the procurement of an independently conducted evaluation.

This evaluation plan was designed as part of a project with the Evaluation Task Force (ETF) seeking to improve and embed evaluation practices across HM Government. The project involves the creation of several demonstrator evaluation plans for five of the programmes within the Government Major Projects Portfolio (GMPP). This evaluation plan falls under the ETF's Information and Communications Technology (ICT) Demonstrators, designed by the externally contracted evaluation partner, PUBLIC, for the evaluation of two *digital* government

programmes.

PUBLIC has worked with His Majesty's Revenue and Customs (HMRC) to produce a detailed evaluation plan for the Single Customer Account (SCA) programme. The SCA is a single destination for individuals and businesses to manage their tax affairs with HMRC. With access to basic information and critical services divided across multiple systems and accounts, the existing HMRC online experience is highly fragmented. The SCA will deliver an omni-channel user experience that makes interacting with the tax system easy while easing demand on non-digital channels.

As of May 2024, the SCA programme is three years into its delivery, with several HMRC services partly onboarded and further phases of development to be delivered through an agile approach. As with many projects within the GMPP, HMRC SCA has a comprehensive business case that details the primary costs and benefits of the programme. A Theory of Change and high-level evaluation plan have also been created during previous phases of evaluation work. Although the SCA is closely aligned with the One Login programme for digital identity verification, and the Unique Customer Record (UCR) for improved data management practices internally to HMRC, these are beyond the scope of this evaluation and outcomes are to be considered separately.

This evaluation plan takes advantage of several features of HMRC and the SCA programme that will be relevant to many other large-scale, centrally-managed, digital projects in central government. This includes making use of HMRC's historical data, access to data through its internal services, and digital methods for improved data collection. The selected methods also consider a key constraint in the lack of natural control groups for the unique services that will be impacted by the SCA programme.

Building on the business case, previous Theory of Change and high-level evaluation plan, this plan includes many of the same outcomes, redefined through specific, monitorable metrics. Finally, we identify several additional outcomes in the short- and long-term to demonstrate a logical, causal flow through to programme impacts.

Section 1 Theory of Change

The HMRC Single Customer Account Theory of Change (ToC) illustrates the causal flow between the inputs, their related activities, and the tangible product outputs; and between the objective outcomes and the intended impacts of the product as a whole.

This ToC acknowledges that the SCA is in ongoing development, with multiple previous, current and future planned phases of delivery for various product functionalities. Although this is unlikely to alter the overall SCA impacts, an evolving product output has potential implications for future additional outcomes and/or currently inconceivable impacts. This ToC is therefore only an analysis of the SCA product in its current state at the time of writing this report, May 2024.

Figure 1 illustrates the HMRC SCA ToC outlined in the above report. Arrows are used to indicate causal relationships. The ToC flows from input over activities, outputs, and outcomes to the impact, which are the overall goals of the SCA programme.

Inputs

There are four key inputs to the SCA programme:

- the funding allocated to HMRC for the build, roll out, run and scale-up of the SCA.
- the HMRC team, including a mix of civil servants and contractors, who are leading and coordinating the development and roll-out of SCA.
- the managed service contracts awarded to specialist suppliers to provide specific deliverables or services that are provisioned more efficiently by commercial partners.
- resources, including analytical software, the IT infrastructure and procurement of tools.

Activities

Key activities include:

- development of the SCA
- HMRC staff recruitment and training
- HMRC staff process review
- customer education and training
- customer research and segmentation
- data management improvements

Development of the SCA

The range of work that has led to the development of the SCA service from a programme, IT build and test, and operational perspective. This includes the strategy and programme delivery, the development of the front-end customer experience, the internal capabilities required for the back-end systems, and the joining up of front and back-ends.

HMRC staff recruitment and training

The recruitment and training of HMRC staff to develop and deliver SCA. Resourcing to ensure HMRC staff are in the right places at the right time. Recruitment activities to ensure the programme can be fully implemented. Training material to ensure staff have the capabilities to implement new processes.

HMRC staff process review

Work undertaken to review data on HMRC staff processes to identify and implement feedback in a continuous loop.

Customer education and training

Individual customer and business education to ensure awareness of new systems, enhancing take-up, including amongst intermediaries and other tax professionals (e.g. agents), and providing customers and businesses with the skills to self-serve. This includes work in the development of clear messaging and direct training material.

Customer research and segmentation

Social research to understand the customer journey, including research into the barriers to customer and business take-up and levels of confidence and competence. Work to understand contact behaviour and barriers to self-service, the digital propensity of different sub-populations and the requirements to ensure internal business systems can be enhanced to work with SCA. This also includes using the Customer Lab to test approaches to customer interaction, customer education and how systems are embedded.

Data management improvements

The combination of multiple ongoing activities to improve HMRC data management practices, including the building of data dashboards for regular reporting on business priorities, and building the customer segmentation to provide a view of the target population, enabling reporting of SCA vs. non-SCA customers.

Outputs

The end-to-end SCA service can be described as the singular output in this ToC. We understand that the SCA programme has been developed and delivered through an agile approach, with the design and release of the service completed at the feature level. However, as no single element of the service exists in isolation, and therefore contributes to any single outcome alone, we will evaluate the impact of the SCA service as a whole, rather than attempting to attribute individual causality to the many features of the SCA.

SCA service

The end-to-end SCA service, the web and mobile apps, their features such as customer registration, the access to different services available within SCA, the tailored customer accounts, the view of customer tax and benefits affairs, and tools such as the Digital

Assistants tool.

Specific features that relate to individual outcomes include:

- The new digital channels for customer contact, now comprising multi-channel customer support across all platforms
- Automated nudges that alert customers when errors are detected while completing forms or providing information
- Self-service functionality that allows customers to access services or complete activities without direct assistance from HMRC staff
- Paperless services that provide customers with outputs from HMRC service interactions digitally

Internal processes

New ways of working and internal processes as a result of training and process review that sit alongside the SCA. This includes new sharing of best practices and the adoption of agile processes.

Specific features that relate to individual outcomes include:

- Controls to detect suspicious activity, making it more difficult for customers to commit fraud
- The automation of previously manual and laborious HMRC staff tasks

Customer insights

Outputs from customer research and segmentation, including a clearer view of the customer journey.

Data management practices

Improved data management practices from matching and merging records across tax heads, general improvements to customer record management, data dashboarding and data segmentation approaches.

Outcomes

Many outcomes of this ToC have originated in the previous iterations of the HMRC Theory of Change. These outcomes follow the same intended metrics as described in the previous SCA Evaluation Plan, including their measures and data sources. To avoid duplication or confusion, some of these original outcomes have been merged or separated into individual outcomes. Finally, several additional outcomes have been defined following the more discursive benefits found in the SCA Business Case, or those that have arisen through workshops with the HMRC SCA team.

The following outcomes are distinguished by their likely time until realisation, as short-, intermediate- and long-term outcomes.

Short-term outcomes

The first short term outcome is that customers use digital contact services rather than phone and post more frequently. As customers can interact with HMRC more easily through online contact services directly within the SCA service, we expect a channel shift from phone and postal contact services to online, digital contact services.

Secondly, customers make fewer errors. Digital ‘nudges’ and ‘prompts’ deflect completion errors and better support customers in meeting their obligations, thus limiting levels of failure to take reasonable care, specifically linked to FTNAE and HICBC errors in the PAYE population. Additionally, customers have a full view of their tax and benefit affairs in one place, reducing errors that may previously have occurred from a lack of customer awareness across multiple HMRC services.

It is also expected that customer difficulty of committing fraud is increased because strengthened controls to detect suspicious activity allows HMRC to intervene before fraud has been committed. Similarly, digitising previous manual interventions and paper correspondence between HMRC and their customers allows for tighter security controls that are less susceptible to malicious or fraudulent activity.

HMRC staff's manual tasks will become increasingly automated because, through digitisation, certain tasks can be automated, simplified and optimised through processes, data scripts and tools. Some lower-level manual tasks can be removed from the HMRC staff workload.

HMRC staff's insights into the customer journey will improve. As a result of customer research and segmentation, clear mapping of the customer journey through the SCA, and the uptake of corresponding staff training activities, HMRC staff have a better understanding and improved insights into the customer journey. It is also expected that customers utilise self-service more frequently as automated self-serve channels provide information and reassurance to customers, and allow customers to perform activities to meet their obligations with HMRC in a compliant way without direct intervention or assistance from HMRC staff.

New functionality within the SCA, combined with the application of insights from customer research activities, improves the general usability of HMRC services. Additionally, customers can contact HMRC using tools such as digital assistants, and utilise other digital tools which are much faster than offline routes. Finally, the number and proportion of services that an intermediary or agent can access on behalf of their client is increased, allowing for multiple routes for customers to access HMRC services.

HMRC data security is likely to improve, as new data management practices include the storage and management of customer data in a secure way across HMRC channels, in coordination with the Unique Customer Record (UCR) programme. Additionally, secure online communications present further data security benefits in relation to phone and post customer correspondence channels. HMRC data accessibility is increased as improved data management practices, including data dashboarding and data segmentation approaches, allow HMRC staff to access data in a more accessible and timely manner.

HMRC data quality is increased due to improved data organisation practices including customer record management and the matching and merging of customer records, which

improves the quality of customer data that HMRC have access to. Also, the digitisation of HMRC services allows for more data to be captured automatically throughout the customer journey, increasing the overall quantity of HMRC data.

Finally, paper outputs from HMRC service interactions are reduced. The use of online HMRC services reduces the need to contact HMRC via post or through the submission of paper forms. Similarly, the digital SCA allows customers to elect to receive paperless outputs from accessing HMRC services.

Table 1: Short-term outcomes and their intended metrics

Number	Short-term outcome	Intended metric
1.01	Customers use digital contact services rather than phone and post more frequently	1. Volume of telephone customer contact service interactions decreases 2. Volume of postal customer contact service interactions decreases 3. Volume of digital customer contact service interactions increases
1.02	Customers make fewer errors	1. Customer submissions requiring staff intervention decrease 2. Customer service tickets raised related to customer error decreases
1.03	Customer difficulty of committing fraud is increased	1. Volume of fraud prevented increases 2. Total tax collected increases 3. Tax gap (*Decrease in collectable tax - collected tax*) decreases
1.04	HMRC staff's manual tasks are automated	1. Qualitative perception of staff workload improves
1.05	HMRC staff's insights into the customer journey are improved	1. Qualitative perception of staff understanding of the customer journey improves 2. Qualitative perception of staff sympathy towards customer errors improves
1.06	Customers utilise self-service more frequently	1. Customer usage of self-service SCA functionality increases 2. HMRC staff manual intervention in areas covered by new SCA self-service functionality decreases
1.07	The usability of services is improved	1. Qualitative perception of customers on the usability of HMRC services improves
1.08	HMRC data security improves	1. Number of HMRC data incidents reported decreases 2. Qualitative perception of customers on HMRC data security improves

Number	Short-term outcome	Intended metric
1.09	HMRC data accessibility is increased	1. Average number of clicks staff are required to make to access required data decreases 2. Qualitative perception of HMRC staff on their access to relevant data in a timely manner improves
1.10	HMRC data quality is increased	1. HMRC measure of data quality improves 2. Estimated number of duplicate records in HMRC data system decreases 3. Qualitative perception of HMRC staff on the quality of the data that they have access to improves
1.11	HMRC data quantity is increased	1. Quantity of data records held by HMRC increases 2. Demographic data collected on HMRC customers increases 3. Number of digital customer interactions increases
1.12	Reduced paper outputs from HMRC service interactions	1. Volume of postal customer contact service interactions decreases 2. Quantity of paper expended in HMRC service operations decreases 3. Volume of staff clicks on “print page” function decreases

Intermediate-term outcomes

One key intermediate outcome is that, as customers reduce their use of slower phone and email contact services, make fewer errors, and make use of self-service SCA functionality, customers require the use of contact services less and for lower total time. There are four metrics associated with this outcome. The first can be calculated by finding the number of customer contact service interactions that require no human intervention (in other words, they are resolved by automated responses) and dividing this by the total number of customer contact service interactions. The fourth metric can be calculated by dividing the number of customer contact service interactions that require no follow-up (i.e. are resolved in a single instance) by the total number of customer contact service interactions.

Another expected outcome is that, as a result of fewer errors and increased difficulty in customers committing fraud, more customers elect to voluntarily comply to ‘get it right’ and avoid the errors which contribute to the Tax Gap.

In the intermediate-term, staff admin burden is reduced, as self-serve channels allow customers to access information and/or perform activities without requiring immediate, direct support or intervention from a member of HMRC staff. It is expected that customer admin burden is also reduced, as automated self-service journeys allow customers to quickly

access and complete HMRC services online. Similarly, with automated nudges and prompts, completion errors are deflected, resulting in fewer resubmissions of information with the requirement of HMRC staff support. Simplification of services from reduced touch points, easier access and improved levels of reassurance reduces the burden on customers.

Inclusivity and accessibility can be improved by reducing the barrier to entry to access HMRC services. Online verifications require less resource investment from end-users. Similarly, digital services may require less travel, for instance, in the case of utilising postal contact channels. The second metric represents the proportion of total complaints which are related to accessibility.

The final intermediate-term outcome is that data accuracy is improved. Through the combination of improved data accessibility and increased data quality and quantity, the data accuracy that HMRC staff access is improved, allowing staff to gain further insights into HMRC customers.

Table 2: Intermediate-term outcomes and their intended metrics

Number	Intermediate-term outcome	Intended metric
2.1	Customers use contact services less frequently	1. Volume of total customer contact service interactions decreases 2. Customer contact service total interaction time decreases 3. Customer ‘containment’ rate increases 4. Customer ‘once and done’ contact rate increases
2.2	Voluntary compliance increases	1. Sustained total tax collected increase 2. Sustained tax gap (decrease in collectable tax - collected tax) decrease 3. Calculation: 4. HMRC Strategic Picture of Risk (SPR) improves 5. Volume of error reported by customers (ISBA) increases
2.3	Staff admin burden is reduced	1. Staff time spent on low-level admin tasks decreases 2. Qualitative perception of staff on the proportion of workload as admin decreases
2.4	Customer admin burden is reduced	1. Total customer time spent on low-level admin tasks decreases 2. Qualitative perception of customers on the level of admin required to access HMRC services decreases

2.5	Accessibility of services is improved	1. Qualitative perception of customers on the accessibility of HMRC services improves 2. Proportion of complaints/customer service tickets raised related to accessibility of HMRC services decreases 3. Uptake of benefits among typically underrepresented target audiences increases
2.6	Data accuracy is improved	1. HMRC measure of data accuracy improves

Long-term outcomes

There are two key long-term outcomes. Firstly, with a sustained reduction in the demand for contact services overall, and reduced demand specifically in more time-consuming channels of customer contact, the demand for HMRC contact services is reduced. Over time this may result in the requirement of less HMRC contact service staff. One metric for this outcome is sustained customer 'containment' rate, which represents the proportion of customer service interactions which are resolved in a single contact. With SCA implementation, it is expected that the containment rate will increase, converging towards a value of 1. Another metric is the sustained customer 'follow-up' rate, which represents the proportion of customer contact service interactions that require follow-up. The follow-up rate is expected to fall with SCA implementation, eventually converging towards 0.

The second long-term outcome is a faster customer journey which is also more convenient and straightforward. It is expected that the admin burden on customers is reduced as they make more frequent use of self-service and automated tasks. Improvements in the usability and accessibility of HMRC services also contribute to an enhanced customer journey, increasing the convenience, ease and timeliness of accessing HMRC services. This outcome can be measured using several metrics, including the sustained customer 'once and done' contact rate, as well as the customer application 'dropout' rate, which represents the proportion of total applications which were started but abandoned before completion and is expected to fall following SCA implementation.

Table 3: Long-term outcomes and their intended metrics

Number	Long-term outcome	Intended metric
3.1	Demand for contact services is reduced	1. Sustained volume of total customer contact service interactions decrease 2. Sustained customer contact service total interaction time decrease 3. Sustained customer 'containment' rate increase 4. Sustained customer 'follow-up' rate decrease 5. Number of FTE/Customer service agents working in telephony roles decreases

Number	Long-term outcome	Intended metric
		6. Number of FTE/Customer service agents working in post demand tasks decreases 7. Total FTE/Customer service agent headcount decreases
3.2	Customer journeys are quicker, more convenient and straightforward	1. Qualitative perception of customers on the customer journey improves 2. Qualitative perception of customers on trust in HMRC improves 3. HMRC Customer Satisfaction (CSAT) scores improve 4. HMRC Customer Net Easy scores improve 5. Total number of customer complaints decreases 6. Sustained customer 'once and done' contact rate increase 7. Sustained customer service tickets raised related to customer error decrease 8. Customer application 'dropout' rate decreases

Impacts

The following ToC impacts capture the longer-term, cumulative, or wider social impacts that are the driving forces behind achieving each outcome of the HMRC SCA programme. Assuming that the outcomes are achieved, the impacts aim to describe why each outcome is ultimately of value for HMRC, the UK Government or the wider UK population. The impacts identified as a result of the SCA programme are:

- HMRC runs more efficiently
- Customer compliance increases
- HMRC staff satisfaction increases
- Customer experience and perceptions of HMRC improve
- Environmental impact of HMRC is reduced

The following subsections outline potential mechanisms through which the outcomes of this project may lead to these impacts.

HMRC runs more efficiently

HMRC reduces the tax gap through fewer customer errors, increased voluntary compliance and increased difficulty of committing fraud. It will reduce staff admin burden through automation of manual low-level tasks, and thereby create a channel shift with contact services from more burdensome phone and post to more efficient digital channels for HMRC services, resulting in a reduction in staff time expenditure in assisting customers, staff costs and ultimately staff headcount.

This impact relates to the critical success factor of improved efficiencies and cost savings for HMRC, and links to HMRC Strategic Objectives SO1, SO2, SO4, SO5.

Customer compliance increases

The overall compliance risk is reduced as customers make fewer errors and increase their voluntary compliance. This is because customers are being nudged onto the correct course of action; find it more difficult to commit fraud with the SCA's automated interventions and more secure processes; and have increased capabilities in organising their tax and benefit affairs.

This impact relates to a critical success factor of the SCA programme in increasing compliance, and links to HMRC Strategic Objectives SO1, SO2, SO3 and SO5.[1]

In the longer term, this increase in customer compliance is expected to result in a reduction in the tax gap (collectable minus collected tax).

HMRC staff satisfaction increases

HMRC staff satisfaction increases as a result of improved HMRC staff insights into the customer journey which lead to a reduction in staff frustration; a reduction in low-level manual tasks through automation leading to reduced staff admin burden; and through new processes and ways of working with improved data accuracy. This is expected to lead to an improvement in staff experience and confidence, staff engagement, and ultimately staff satisfaction.

This impact relates to a critical success factor for the SCA programme in delivering a positive user experience for colleagues using HMRC digital channels, and links to the HMRC Strategic Objective SO4, making HMRC a great place to work.

In the longer term, this increase in staff satisfaction is expected to result in improvements to staff Great Place to Work surveys and a reduction in staff churn.

Customer experience and perceptions of HMRC improve

Customer experience and perceptions of HMRC improve as a result of reduced customer error and admin burden and improved customer journeys with the ability to access all HMRC services in one place, in a more convenient, easy, and timely way. Customer experience will also improve because of the increased usability and accessibility of services, and the ability to contact HMRC more easily. Similarly, customer perceptions of HMRC improve through increased data security, reduced fraud, improved data accuracy held by HMRC staff and modernised digital experience.

This impact related to the critical success factors of the SCA programme of creating a simple and easy user experience, making the HMRC digital experience the first choice for everyone in the UK to engage and manage their tax affairs, and finally, links to the HMRC Strategic Objectives SO1, SO2, SO3, SO5.

In the longer term, this improvement in customer experience and perceptions of HMRC should result in improved customer trust in HMRC, improved HMRC reputation, and

[1] HMRC's annual report and accounts 2023 to 2024: performance overview (2024): Available at: https://www.gov.uk/government/publications/hmrc-annual-report-and-accounts-2023-to-2024/hmrcs-annual-report-and-accounts-2023-to-2024-performance-overview

ultimately further uptake of HMRC services.

Environmental impact of HMRC is reduced

As more HMRC customers elect to receive paperless outputs from accessing HMRC services, reducing paper usage across HMRC services over time, the overall environmental impact of HMRC is reduced. As with any new digital service, channel-shift benefits can drive improved environmental impacts, especially when compared to non-digital services.

Theory of Change discussion

In this section, we discuss any assumptions made through the logic of the ToC above and interrogate the dependencies for the successful implementation of the SCA. We also explore the potential unintended consequences of the implementation and ongoing delivery of the SCA programme.

External dependencies

The successful implementation of the SCA is contingent upon several external dependencies that may influence the outcomes of the intervention. These external dependencies are acknowledged as Caveats to Benefits Realisation in the SCA Business Case, and are important to highlight here for future evaluators of the SCA service to take into consideration. These are:

- capacity and capability
- operating model
- analytics
- policy and legislation

Sufficient capacity and capability is required in order to achieve the recommended development of the SCA programme. Thus transformation activity will need to be accelerated, with IT partners' full support to deliver and exploit the full opportunity to realise the benefits potential over a shorter period.

The operating model will need to support the digital customer interaction ambition, with appropriate customer service options as part of the channel strategy. Also, the business will need to be able to accept the volume of change required to exploit the full opportunity.

There is a need for analytics to be able to understand customer contact and track across the channels offered, to influence customer behaviour in more detail and to identify cross-channel repeat contact and pain points for our customers.

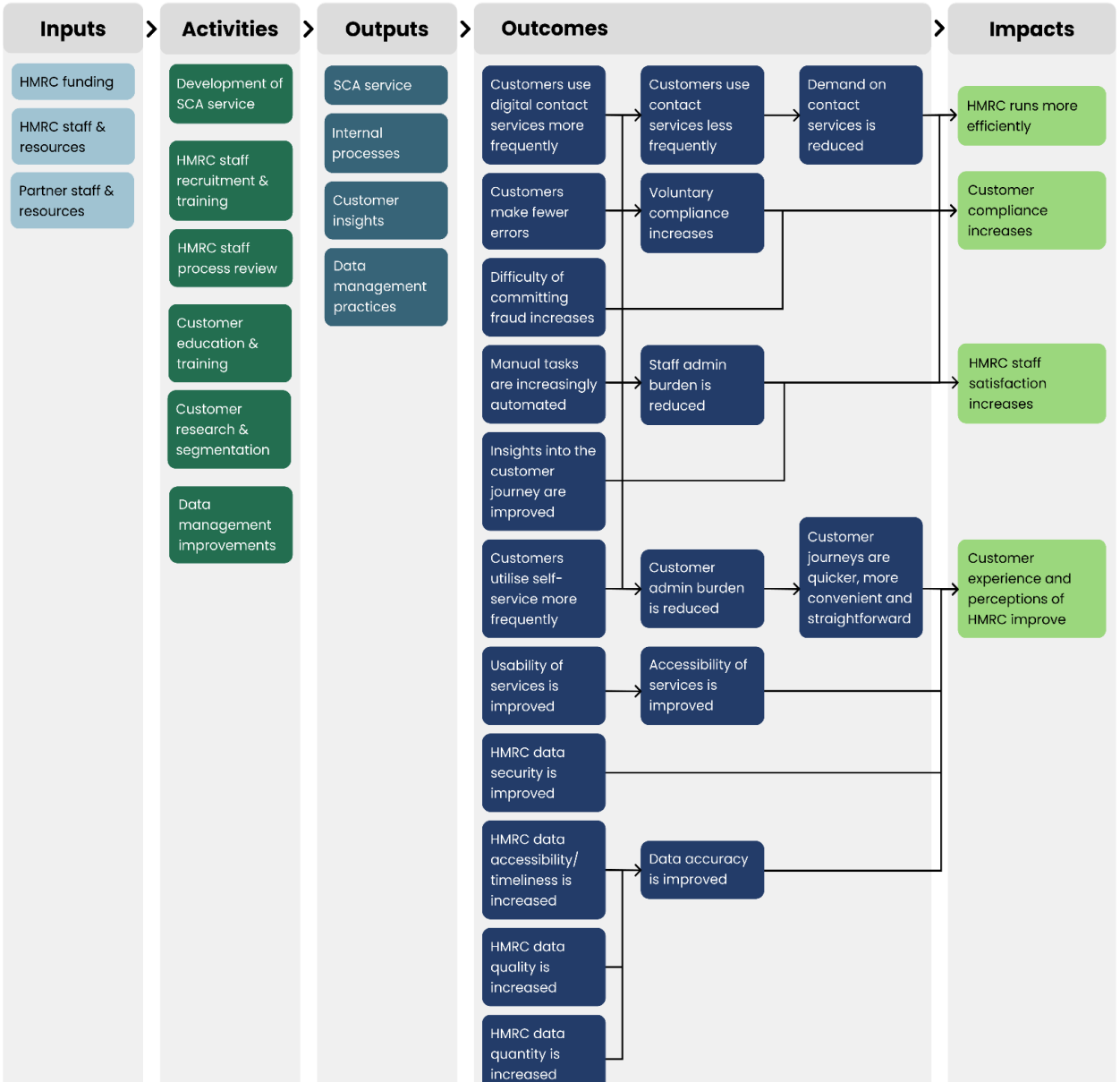
Unintended consequences

There may be a digital divide in access to online services. As described in Outcomes 1.07 and 2.5, new functionality such as digital assistants or prompts that reduce error, both simplify customer journeys and reduce the customer admin burden by improving the usability and accessibility of HMRC services. However, as these functionalities are moved or built into the online SCA, their corresponding outcomes are only felt by users who can access the internet.

There is also an environmental cost of running a digital service. As described in Outcome 1.12, the operation of HMRC service through a central digital account results in a possible reduction in the use of paper, through reduced paper outputs and customer correspondence. This avoided use of material has a positive environmental impact, however, there is an environmental cost associated with running a digital service which must be assessed to understand the net environmental outcome.

Finally, there may be reduced trust from Government verification systems centralisation and data aggregation. As described in Impact 4 (the reduction of HMRC's environmental impact), increases in data security and accuracy, reduced fraud, and a modernised digital experience improve customer perceptions of HMRC and over time, can increase trust in HMRC and its services. However, increased data tracking, data aggregation and centralisation of services can also lead to negative perceptions and a reduction in trust, and therefore must be taken into account when understanding the net change of customer perceptions and trust in HMRC and its service.

Figure 1: Single Customer Account Theory of Change Visual Logic Model



Section 2 Process evaluation outline

In this section, we set out the considerations for a process evaluation of the SCA programme in the two strands of development and delivery. This approach concerns the SCA in its current state, with many HMRC services onboarded to the service and planned continued onboarding of additional services and added functionality.

Following Magenta Book best practices, this process evaluation gathers lessons learned from the delivery of SCA, from its inputs and activities in building the outputs.[2] Broadly, the questions that are asked seek to identify:

- What worked well and less well, and why?
- What could be improved?
- How has the context influenced delivery?

More specifically, this process evaluation outline seeks to evaluate the extent to which the SCA has followed good software development practices, and whether the services have been built in an efficient and cost-effective manner, in line with practices outlined in the GDS Service Assessment and the Technology Code of Practice[3].

Development

The aim of understanding the development journey of the principal platform or its updates is to learn from successes and challenges. This allows HMRC to improve similar future processes, identify risks and mitigation strategies and build institutional knowledge. To arrive at this knowledge base, we propose the following guiding questions.

1. Was the development process efficient?

This could be answered by examining whether timelines were adhered to, whether resources were effectively allocated, and whether there were any bottlenecks or delays. In particular, this question might focus on processes and teaming structures established with delivery partners.

This is best answered by reviewing programme documentation. Interviews with members of the development team may be conducted to provide clarification regarding any gaps that are identified.

2. Was the project effectively managed?

In answering this question, we determine whether project goals were clearly defined, if communication channels were effective and if risks were adequately managed.

This is best answered through in-depth interviews with members of an output development team.

---

[2] HMT, The Magenta Book (2020). Available at: https://www.gov.uk/government/publications/the-magenta-book

[3] Central Digital and Data Office, The Technology Code of Practice (2023). Available at: https://www.gov.uk/guidance/the-technology-code-of-practice

3. How were stakeholders engaged and for what?

This requires evaluating how stakeholders were involved throughout the development process. We can assess whether their input was solicited, how feedback was incorporated, and whether there were any challenges in stakeholder management.

This is best answered through in-depth interviews with stakeholders and project managers.

4. Did the development process follow the principles of user-centred design, outlined in the Service Manual and Technology Code of Practice?

This includes evaluating the usability testing conducted, user research methods employed, and how user feedback influenced design decisions.

This is best answered by reviewing programme documentation. Interviews with members of the development team may be conducted to provide clarification regarding any gaps that are identified.

5. Were there any technical challenges in development and how were they addressed?

Identify any technical challenges encountered during development and assess how they were addressed. This could include issues related to scalability, security, interoperability, or integration with existing systems.

If technical challenges and their resolutions were routinely documented during the development process, a programme documentation review would be sufficient.

6. Which standards and regulations were adhered to and to what extent?

Evaluate whether the SCA development adhered to relevant standards and regulations, such as accessibility standards, GDPR and other information governance standards (e.g. ISO and Cyber Essentials), and the Technology Code of Practice.

This data can be collected through a review of the programme’s documentation.

7. How thorough was the quality assurance testing?

This entails an assessment of the effectiveness of quality assurance processes and testing methodologies employed during development. It includes examining whether adequate testing was conducted to ensure the platform's functionality, security, and performance.

Surveys can be used to identify which aspects of each output have an established quality assurance process. Where processes exist, programme documentation can be used for assessment. Where processes are missing, interviews with relevant team members can be conducted to provide context and reasoning.

8. What is the quality of the project's documentation?

This requires evaluating the completeness and usefulness of documentation produced during the development process. It includes technical documentation, user manuals, and training materials.

This can be done through a review of the programme's documentation.

Delivery

The aim of assessing the delivery of the SCA to end-users is to enable continuous improvement. It is therefore recommended that the delivery of the outputs to end-users is routinely evaluated. To best enable continuous improvement, this arm of the process evaluation should be done as frequently as budget permits.

1. Was the output efficiently delivered to the end-user?

This involves identifying the time it takes for services to integrate with the new platform, the ease of use for end-users, and any bottlenecks or inefficiencies in the process.

This data can be collected through programme documentation as well as performance data from the back end of the SCA system.

2. How was quality controlled?

This involves identifying where quality assurance checks already exist and where they could be added. For existing processes, this would involve assessing the robustness of the checks.

This data can be collected through surveys and interviews with relevant project delivery teams.

3. What was the end-user's experience?

This requires understanding user pain points and preferences. The usability of the output can also be assessed.

This data can be collected through surveys and focus groups with end-users.

4. How do current protocols ensure both security and compliance standards are met throughout the delivery of this digital public service?

This involves assessing data protection measures, and authentication protocols, including adherence to GDPR and other information security requirements. It also identifies gaps in such checks for protection.

This data can be collected through programme documentation as well as performance data from the back end of the SCA system.

Detailed instructions for carrying out each of the methods described in this process evaluation outline, including in-depth interviews, surveys and questionnaires, focus groups and programme documentation

review, can be found in Annex A: Process Evaluation Methods.

Section 3 Impact evaluation outline

In this section, we discuss the opportunities for, and constraints to, evaluation that inform the selection of our proposed approaches, including A/B testing and synthetic control. We provide an overview of each of these methods, followed by a plan that outlines the key considerations and steps required to implement each approach.

The impact evaluation outlined in this section specifically considers the evaluation of the outcomes described in the ToC. These outcomes are observed, or 'monitored', via a series of metrics. Each outcome has been assigned metrics in the ToC Metrics database, which we propose are monitored by the evaluation team. However, not all monitored metrics in this database are to be evaluated, with specific exceptions mentioned where attribution of causality is not feasible.

Opportunities for evaluation

The impact evaluation plan outlined in this section takes advantage of several features of large-scale, government-led, digital projects. These features present unique opportunities to reduce the burden on the evaluator, and ultimately, minimise the cost of conducting the evaluation. We recommend the utilisation of HMRC historic data, data from internal HMRC services, SCA performance data, and where possible, opportunities for embedded feedback collection from service users.

As a large, long-standing, central government department, HMRC has large amounts of historical data across many of the metrics identified in this evaluation plan. Data of this quantity, collected for this duration, opens up the possibility of synthetically constructing a control group through the estimation of changes to the metric in the absence of the treatment effect. This method of evaluation, known as *synthetic control,* is explored further in the selected evaluation methods subsection below for metrics where adequate data availability has been suggested by the HMRC SCA team.

The onboarding of services to the SCA is to be conducted via an agile approach, which allows for the possibility of randomisation and enables an experimental impact evaluation.

Performance data in digital projects refers to the data that may be passively and continuously collected via the digital systems employed, or naturally collected through the normal running of the service. This typically includes metrics related to user engagement statistics, website/service traffic, load times/technical performance, and error rates. We have identified a wide range of possible metrics to be collected via performance data, including customer submissions requiring staff intervention, customer usage of self-service SCA functionality, and the total number of customer complaints. Additional sources of performance data may be identified as the SCA system is further developed, and we encourage the future evaluation team to consider these data sources and prioritise these metrics for a cost-effective evaluation.

Embedded feedback collection leverages the digital format to integrate user feedback mechanisms directly into the system interface. This approach enables the collection of immediate and context-specific insights from users while they interact with the digital environment. Methods such as pop-up surveys, feedback buttons, and interactive chatbots can be employed to gather qualitative and quantitative feedback. This real-time data reflects user experiences and satisfaction levels without the typical delay

associated with traditional feedback methods. Additionally, embedded feedback collection negates the requirement for outreach and communication with users post-experience with the system, reducing the time taken and cost to collect the data.

Constraints to evaluation

In addition to the opportunities outlined above, there are aspects of HMRC and the SCA programme that present a unique set of challenges, ruling out the use of specific evaluation approaches. The primary constraint here is in the widespread rollout to HMRC customers and the lack of natural control groups for HMRC's specific services. This is because the SCA service is to be made accessible to all HMRC customers simultaneously. The broad implementation of the SCA to all eligible taxpayers means that there is no segment of the relevant population unaffected by the service, which prevents us from evaluating the impact of the programme using an experimental evaluation design.

Nonetheless, applying a methodological approach aimed at understanding the cause-and-effect relationships between the SCA and its expected outcomes provides useful insights into the benefits of this service. The absence of a natural control group - a segment of the population unaffected by the SCA - necessitates creative and alternative approaches to establish causality. Moreover, the intricacy of the SCA programme's impact across different taxpayer services adds another layer of complexity to its evaluation. To address these challenges, combining various methods of causal inference can provide a more robust understanding of the SCA programme's impacts. This blended approach can compensate for the limitations of any single method, offering a fuller picture of how the programme works and for whom it works best.

Selected evaluation methods

In this subsection, we discuss our selected evaluation methods for the impact evaluation of the SCA, based on the opportunities and constraints discussed above. For each, we detail the general design, requirements, risks and limitations.

Where SCA programme delivery and the planned rollout of features allow, we propose A/B testing. In this experimental method, we observe and compare the performance of two versions of the service or its features. If sufficient historical data on the evaluation metric is available, or comparable units not affected by the intervention are identified, we propose synthetic controls. In this quasi-experimental method, we artificially construct an evaluation control group based on the data available.

For instances where sufficient data is not available, we suggest identifying an instrumental variable or conducting propensity score matching. Finally, for more user-centred and experiential outcomes, or where experimental or quasi-experimental methods are not feasible, we propose qualitative methods.

Our proposal can be summarised as follows:

- use A/B testing as the first-best evaluation strategy where programme delivery and the planned rollout of features allow
- use synthetic controls as a second-best strategy where sufficient historical data or comparable groups are identified

- use an instrumental variable or propensity score matching as a third-best strategy where the data requirements for a synthetic control are not met
- use qualitative methods as a fourth-best strategy for evaluation where A/B testing or synthetic controls are not feasible, or where a qualitative approach is most appropriate for the specific metric

Experimental methods

An experimental method of causal inference is a research design that seeks to establish a cause-and-effect relationship by constructing a control group using randomisation. This method is considered the gold standard for causal inference because it allows for the direct observation of how changes in one aspect directly influence another, minimising the influence of external factors.

A/B testing

A/B testing, a standard in software development, operates on the same principles as experimental methods of causal inference. In an A/B test, two versions (A and B) of a product or service are compared, where one serves as the treatment and the other as the control. This method allows for precise measurement of the effect of specific changes or features on user behaviour or outcomes. The random assignment of participants to either the treatment or control group helps ensure that the results are due to the variable being tested rather than other external factors.

Testing at the feature level involves evaluating specific modifications or additions to a component of the programme, such as a new user interface element or an updated process. However, when feature-level changes are expected to significantly affect desired outcomes, these insights can inform broader programme-level evaluations. Such evaluations assess the collective impact of multiple features or comprehensive changes to the entire programme, providing a holistic view of effectiveness and user experience. By strategically linking feature-level testing with programme-level outcomes, HMRC can ensure that each modification not only serves its immediate purpose but also contributes optimally to the overarching goals of the SCA.

For example, conducting A/B tests on various digital nudges could directly identify their effects on metrics such as Metric 1.02.1, “Customer submissions requiring departmental intervention decrease", or Metric 1.02.2, “Customer service tickets raised related to customer error decreases."

This approach enables a precise evaluation of how individual elements of the SCA affect taxpayer behaviour and compliance, offering a robust method to refine the platform in alignment with SCA's strategic goals.

By integrating experimental methods to test these and other metrics outlined in the SCA ToC, HMRC can systematically enhance the effectiveness of the SCA, ensuring it meets its objectives of improving compliance, reducing errors, and elevating taxpayer satisfaction through an evidence-based understanding of cause and effect.

Quasi-experimental methods

Synthetic controls

Synthetic control is a technique used to understand the impact of a new policy where a treatment has been applied universally as is the case with the SCA where the service is made available for all eligible taxpayers.

A synthetic control creates a predictive model, of an outcome of interest, that is trained and tested on historical data. This model is then used to project the evolution of that outcome after the introduction of the SCA. As the model's training does not include information regarding this new system of interacting with HMRC, this projection is used as the control observation.

In general, a synthetic control is constructed using historical data on the treated unit as well as potential control groups. These control groups do not need to meet the requirements of other procedures of causal inference as they are used for predictive purposes.

Figure 2: Synthetic Control General Design

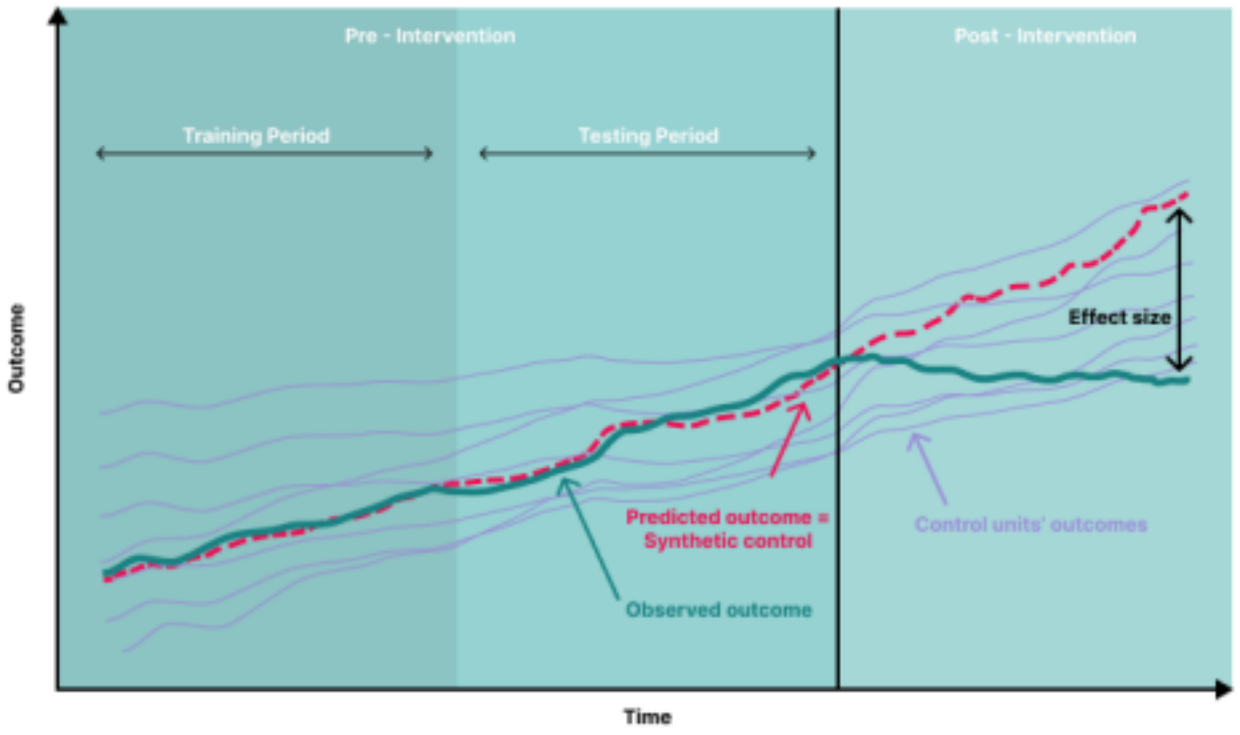


Figure 2 illustrates the general setup for this method. In the Pre-Intervention period, a Synthetic Control is constructed using data on the outcome of interest from control units. This construction follows a weighting procedure, where the size of a weight reflects the predictive power of that control unit.

To establish the weights, the model is developed using pre-intervention data from a training period. The

efficacy of the resulting model is then checked in a testing period. Once a sufficiently predictive model has been developed, the control units' post-intervention outcomes are used to predict the unobserved counterfactual of the treated unit. The difference between the predicted outcome and the observed outcome of the control unit gives the effect size.

Constructing a synthetic control

This section outlines the general approach to constructing a synthetic control and illustrates these principles with a specific example from HMRC's Metric 1.02, which focuses on reducing customer errors through digital nudges.

Data compilation

The first step involves compiling a comprehensive dataset that includes data on both the treatment group (affected by the intervention) and potential control units (unaffected by the intervention). This data should detail not only the primary outcomes but also relevant variables that can influence those outcomes.

For reducing customer errors (Metric 1.02), historical error rates from HMRC databases are gathered, encompassing a significant period before the SCA's introduction. This includes data on types of tax forms, user demographics, and prior interactions with HMRC digital platforms. Error rates are gathered from all HMRC digital services for which this data is available. Services to which the Single Customer Account is not applied serve as potential control units.

Variable selection and methodology

The selection of predictors is crucial for matching the treatment unit with control units. Variables are chosen based on their influence on the outcome and their availability before the intervention.

For Metric 1.02, variables such as the complexity of tax issues, frequency of user interaction with tax platforms, and legislative changes impacting taxpayer behaviour are selected. These variables are used in a statistical model to estimate their weights accurately.

Synthetic control construction

Using the selected data, a synthetic control is constructed. This involves an algorithmic approach to calculate optimal combinations of weights for the selected variables, creating a control group that mirrors the pre-intervention characteristics of the treatment group.

For Metric 1.02, the synthetic taxpayer group is constructed to mirror the error rates that would likely have been observed without the introduction of digital nudges. The effectiveness of these nudges is then evaluated by comparing them against this synthetic baseline.

Evaluation and analysis

Once the synthetic control is established, it serves as a baseline to compare against the actual post-intervention outcomes. This comparison helps isolate the effect of the intervention from other external factors.

For Metric 1.02, the observed post-intervention error rates are compared to those predicted by the synthetic model. Significant reductions in errors attributable to the SCA's digital nudges can be robustly claimed as a direct result of these interventions.

Requirements

In order to conduct the impact evaluation, there are some key requirements that must be met, namely:

- availability of data
- identification of comparable units

Firstly, there must be sufficient historical data on the group affected by the intervention and potential control units. Typically, the more historical data available, the better the synthetic control model will perform, as it allows for a more accurate estimation of pre-intervention trends. A minimum of 2-3 years of data is often recommended to capture enough variability and seasonality in the data, though this can vary based on the frequency of data collection (monthly, quarterly, etc.) and the specific dynamics of the variables involved. This data should also cover a period long enough before the intervention to establish baseline trends.

Secondly, there must be entities (such as regions, organisations, or groups) that were not affected by the intervention but are similar to the treated unit in key characteristics relevant to the outcome being studied. In the case of the SCA, where the change in the customer account is applied universally, potential control units can include the outcome data from users of those benefits or tax accounts that have not yet been moved to the new system. While generally, these are not natural control groups, as the eligibility criteria for entitlements and taxation are systematic, in this setting they may still be used. This is because a synthetic control is similar to a predictive model and so data that can be shown to perform well at predicting an outcome, rather than causing an outcome, is fit for inclusion.

Risks

There are two key risks, the first of which is data mismatch. The success of synthetic control hinges on the availability of comparable data. If the control units don't closely mirror the treatment group before the intervention, the synthetic control may not accurately represent the counterfactual scenario.

The second risk is external influences. Significant events unrelated to the intervention that affect the treatment group can skew the results, leading to incorrect conclusions about the intervention’s effectiveness.

Limitations

Key limitations to the impact evaluation are:

- data availability
- complexity in finding comparables
- interpretation and generalisation difficulties

Firstly, the method's feasibility is directly tied to the extent and quality of available data. In addition to requiring ample historical data on the outcome of interest, it is also necessary to have equally detailed

data on the variables being used to construct the synthetic control. In addition, in many real-world scenarios, finding enough similar cases to build a valid synthetic control can be challenging. Finally, the findings from a synthetic control analysis are highly specific to the context and data used to generate them. This specificity can limit the ability to generalise results to other settings or populations.

Despite these challenges, synthetic control offers a robust framework for evaluating the impacts of large-scale interventions when traditional experimental approaches are not feasible. With careful consideration of these risks and limitations, researchers can leverage synthetic control to gain insightful analyses into the causal effects of programmes like the SCA.

With all of these considerations in mind, we recommend that HMRC considers outcomes 2.1 and 2.2 as the best candidates for a synthetic control approach, when considering different evaluation methods. This is because they represent sufficiently important metrics for the programme that warrant an econometric approach, but also because they can be feasibly evaluated via a synthetic control.

We summarise some potential control units for these outcomes below:

For Metric 2.1, “Customers use contact services less frequently”, potential control units for a synthetic control approach could include:

- data on contact with HMRC regarding taxes or benefits that have not integrated the SCA
- data from other departments that have customer contact service interactions

The data on contact with HMRC do not need to be services that have similar customer service numbers to SCA services, but rather any services where HMRC is able to access historic, reliable data about customer service interactions, with multiple observations. Other departments that could provide data on customer contact service interactions include the Department for Work and Pensions (DWP). Using this as a control unit would involve comparing with services within the DWP that have not undergone digital transformation, as well as alternatives such as NHS Helplines, or Local Council customer service transactions.

For Metric 2.2, “Voluntary compliance increases” potential control units for a synthetic control approach could include tax gap data from other National Tax Authorities not using a service similar to the SCA (a digital end-to-end customer account platform). This would require accessing international tax outcome data, especially tax-gap levels. The OECD’s Tax Administration Database provides a solid basis that could be used for making international tax compliance comparisons.[4]

Instrumental variables

In evaluating the impact of the SCA, instrumental variables (IV) can also provide a sophisticated statistical method to infer causality in scenarios where randomised controlled trials are not practical. IV methods are used to resolve issues of endogeneity—where variables influencing both the treatment and outcome could skew results—by using an instrument that predicts the treatment but is otherwise unrelated to the outcome variables. This allows for the isolation and quantification of the direct effects of the treatment (in this case, SCA usage) on outcomes such as compliance and efficiency.

[4] OECD, Tax Administration Series Database. Available at: https://www.oecd.org/en/data/datasets/tax-administration-series-database.html

Identification of instruments

The selection of appropriate instrumental variables is crucial. An instrument must influence the adoption of SCA but must not be directly related to the outcomes of interest, except through its effect on SCA usage. Examples may include external regulatory changes or economic events that affect SCA uptake but do not have a direct impact on outcome variables like compliance rates or efficiency measures.

The following process can help identify an appropriate instrument:

1. Exploratory data analysis
2. Theoretical justification
3. Correlation analysis
4. Exclusion restriction testing

Initially, an extensive exploration of available data is conducted to identify potential variables that influence SCA usage but are presumed not to have a direct effect on the outcomes of interest.

Then, each potential instrument must be theoretically justified to ensure it affects the treatment (SCA adoption) exclusively through its correlation with the independent variable and not through other channels affecting the outcome.

The third step involves conducting statistical tests to verify that potential instruments are strongly correlated with SCA usage. This often involves regression analyses to demonstrate the predictive power of the instrument over SCA adoption.

Finally, rigorous testing is required to confirm that the instruments do not have a direct relationship with the outcome variables outside of their relationship through SCA usage. This typically involves statistical tests and logical arguments to establish that the potential outcomes are not affected by the instruments directly

Challenges and considerations

Some challenges and considerations to be made when employing IV methods are:

- selection of strong instruments
- risk of biased estimations
- tailoring instruments for each SCA outcome

The difficulty in identifying strong and valid instruments that satisfy both the relevance and exclusion restrictions is a significant challenge. In addition to this, If the instruments are weakly correlated with SCA usage or indirectly affect the outcomes, the estimations may be biased, leading to incorrect conclusions.

Finally, it is important to understand that one instrument cannot universally apply to all outcomes due to varying underlying dynamics specific to each outcome. Each outcome may be affected by different factors, and thus, requires the identification and validation of a unique instrument. This process of selecting and validating instruments needs to be repeated for each outcome chosen for evaluation to ensure the accuracy and reliability of the causal inferences made.

Examples of potential instruments

Some potential options for instruments include:

- regulatory changes in tax filing requirements
- introduction of new digital infrastructure in specific regions
- changes in technology grants for small businesses
- specific campaigns promoting digital literacy among taxpayers
- launch of Complementary Digital Services by HMRC

See Table 4 for more detail on these instruments, along with the specific metrics they could help evaluate and the rationale for instrument choice.

Table 4: Examples of potential instruments for different metrics

Metric	Potential Instrument	Rationale
Increase in digital tax filings and submissions (Metric 1.02)	Regulatory changes in tax filing requirements	Regulatory changes that mandate or encourage digital filing can serve as an instrument for SCA usage, presuming these changes promote the use of digital platforms without affecting the accuracy or compliance outcomes directly.
Regional variations in the adoption of SCA services (Metric 1.04)	Introduction of new digital infrastructure in specific regions	The deployment of improved digital infrastructure, such as enhanced internet connectivity in specific areas, could increase the likelihood of using SCA. This would only influence outcomes through increased or facilitated SCA usage.
Business engagement with digital tax services (Metric 1.06)	Changes in technology grants for small businesses	Offering or modifying technology grants for small businesses to adopt digital tools may serve as an instrument by affecting how readily these businesses engage with digital services like the SCA, without directly impacting their tax compliance or error rates.
Usage rates of digital assistance features within SCA (Metric 1.07)	Specific Campaigns Promoting Digital Literacy Among Taxpayers	Campaigns aimed at increasing digital literacy could indirectly increase the usage of SCA by making users more comfortable with online services. This influence is presumed not to directly affect the outcome metrics like user satisfaction or compliance, only through increased adoption and usage of SCA.
Integration and comprehensive usage of SCA features (Metric 1.09)	Launch of Complementary Digital Services by HMRC	The introduction of additional digital services that complement the SCA could increase overall SCA usage as users integrate these services into their routines. These complementary services should not directly affect specific SCA outcomes like error rates or compliance levels, except through their impact on increasing SCA usage.

Propensity score matching

Propensity Score Matching (PSM) is a statistical technique used to estimate the effect of an intervention by accounting for covariates that predict receiving the treatment. In the context of evaluating the SCA, PSM allows us to compare users who are similar in all observed respects except for their participation in the SCA. This method helps to simulate a randomised control trial by matching each treated unit (SCA user) with one or more non-treated units (non-users or different levels of usage) based on their likelihood of using the SCA, as estimated by a propensity score.

The general application of PSM requires the following phases:

- score calculation
- matching process
- outcome analysis

First a propensity score is calculated for each individual in the study. This score represents the probability that an individual will use the SCA, based on various observed characteristics such as prior digital engagement, demographics, and other relevant factors. The score is typically calculated using logistic regression, where the dependent variable is SCA usage and the independent variables are the factors expected to influence this decision.

Once the propensity scores are calculated, the matching process begins. This involves pairing each SCA user with one or more non-users who have similar propensity scores. The goal is to match users and non-users who are as similar as possible in terms of their calculated propensity to use the SCA, thus creating a control group that closely resembles the treatment group in every respect except the treatment itself.

After the matching is complete, the final step is to compare the outcomes between the matched SCA users and non-users. This comparison can include a variety of metrics, such as compliance rates, error rates, and the overall intensity of SCA usage. The analysis aims to identify any significant differences in these outcomes, attributing them to the effect of the SCA. This step helps to isolate the impact of the SCA from other variables that were accounted for in the propensity score.

Control group

In Propensity Score Matching, a control group is composed of individuals who did not receive the intervention—in this case, those who have not used the SCA or have used it less frequently. These control units are selected to closely mirror the characteristics of the treated units (SCA users) based on their propensity scores. The propensity score reflects each individual's probability of participating in the SCA, calculated from relevant covariates such as demographics, previous digital engagement, and other factors predictive of SCA usage.

By matching treated individuals with control individuals who have similar propensity scores but did not receive the treatment, PSM creates a plausible control group. This group serves as a counterfactual, indicating what the outcomes for the treated group might have been had they not received the intervention. The quality of the control group is critical, as it must ensure that the only systematic difference between the treated and control groups is the intervention itself, thus mimicking the conditions

of an experiment.

Challenges and considerations

Some challenges and considerations to be made when employing PSM are:

- quality of matching
- selection of covariates
- common support problem

The success of PSM heavily depends on the ability to accurately calculate propensity scores and match units effectively, as poor matches can lead to biased results. Also, the choice and availability of covariates to include in the propensity score calculation are critical. Omitted variable bias can occur if important predictors of SCA usage are not available or overlooked. Finally, ensuring that there is sufficient overlap in the propensity scores of the treatment and control groups (common support) is necessary to validly compare the outcomes across these groups.

Qualitative methods

Strictly experimental or quasi-experimental methods may not be feasible or appropriate for every outcome or impact outlined in the ToC. Metrics relating to these outcomes have been highlighted to utilise qualitative approaches to evaluation. Qualitative methods can be particularly useful to capture how the SCA has impacted end users from their own perspective and is, therefore, our suggested evaluation approach for many user-centred metrics. However, a qualitative approach can be taken for the evaluation of a range of metrics in the ToC if an experimental or quasi-experimental method is chosen not to be implemented.

Qualitative methods do not employ treatment and control groups that are typical of experimental or quasi-experimental designs, instead gathering data directly from users, with a focus on experiences, perceptions, and behaviours in a natural setting. Qualitative methods, therefore, prioritise deep insights into context and meaning, rather than determining causal relationships. While qualitative methods can infer theoretical causality by identifying patterns and relationships that suggest possible causal links, they cannot be used to definitively attribute scientific causality to an intervention in the absence of controlled comparisons.

Qualitative methods, therefore, provide a lower level of evidence for the impact of the SCA than the experimental and quasi-experimental methods discussed in this section. However, due to the relative ease of their implementation, qualitative methods can be effective in gaining rapid insights and should form the minimum level of evaluation employed by a future evaluation team. Additionally, we understand that HMRC already conducts social research alongside SCA development and delivery, and we therefore encourage future evaluators to align qualitative evaluation efforts with the existing user engagements conducted as part of HMRC's social research offer.

To assess the impact of a qualitative nature, focus groups or in-depth interviews can be employed. For larger-scale qualitative evaluation, surveys and embedded feedback forms can be effective.

Implementation guide

The implementation guide for the impact evaluation of the HMRC SCA is designed to capitalise on the availability of existing data, which serves as a significant advantage. This data, already collected and stored within HMRC's various systems, provides a rich foundation for comprehensive analysis. While this implementation guide focuses on the delivery of an evaluation using quasi-experimental methods, we strongly suggest considering the use of experimental methods whenever feasible.

As the first-best method for evaluation, we suggest implementing A/B testing alongside ongoing efforts at the feature level, which would require slowing the delivery of the SCA rollout to allow for greater periods of testing. Different A/B test groups can represent different treatment groups, allowing the evaluation team to make A/B test comparisons between treatment groups, but also against our synthetic control groups, described later in this implementation guide.

As a second-best method for evaluation, we recommend the creation of synthetic controls where sufficient data on the individual metrics is available. The first step involves centralising and transforming this data into a coherent format suitable for detailed evaluation. This process will ensure that all relevant information is standardised, ensuring consistency and accuracy in the analysis to follow.

A major challenge in this evaluation is the absence of a natural control group due to the universal implementation of the SCA across all eligible taxpayers. To address this, the evaluation plan includes the selection of innovative causal inference methodologies that can simulate control scenarios. These methods include synthetic control and other quasi-experimental designs that help approximate the effects of the SCA by constructing a comparative framework against what outcomes would have been in the absence of the SCA. This blend of methodologies aims to overcome the limitations of traditional approaches and provides a robust understanding of the SCA's impacts.

The actual implementation of the evaluation will proceed in phases, each targeting different temporal outcomes associated with the SCA. Initially, short-term impacts will be assessed to understand immediate effects and adjustments post-implementation. This is followed by medium-term and long-term analyses, which will look at sustained behavioural changes and the overall efficacy of the system in improving taxpayer compliance and satisfaction. Each phase will conclude with detailed reporting, providing insights into the effectiveness of the SCA and highlighting areas for further enhancement.

Throughout the evaluation process, there will be ongoing reporting and documentation. This ensures that each stage of the evaluation is transparent and accountable. Reports generated after each phase will not only detail findings but also discuss the methodologies used and the rationale behind them. This approach facilitates continuous feedback and allows for adjustments to be made in real-time, ensuring the evaluation remains adaptive and responsive to observed dynamics.

Finally, stakeholder engagement is an integral part of the evaluation process. Regular updates and consultations with relevant HMRC departments and external experts will help validate the findings and ensure the evaluation's conclusions are well-supported and actionable. This collaborative approach will aid in refining the SCA based on empirical evidence, ultimately enhancing the service's value to both HMRC and the taxpayers it serves.

To visualise the process of conducting each of the phases described above in a single evaluation, we

have produced the Gantt Chart in Figures 3, 4 and 5. In these charts, each phase is sequential, rather than simultaneous.

Finally, we recommend that the future evaluation team use qualitative methods for evaluation where A/B testing is not feasible, sufficient data is not available, or where direct user engagement is found to be the most appropriate approach. Where applicable, the evaluator should employ one-to-one interviews, focus groups, surveys and embedded feedback forms to collect qualitative data. A/B testing forums can also provide useful qualitative input, and we encourage the future evaluation team to participate in ongoing A/B testing efforts.

Detailed instructions for carrying out the methods described in this impact evaluation outline, including instrumental variables, and propensity score matching, can be found in Annex B: Impact Evaluation Methods.

Figure 3: Phase 1 and 2 proposed quasi-experimental impact evaluation plan Gantt Chart

Figure 4: Phase 3 proposed quasi-experimental impact evaluation plan Gantt Chart

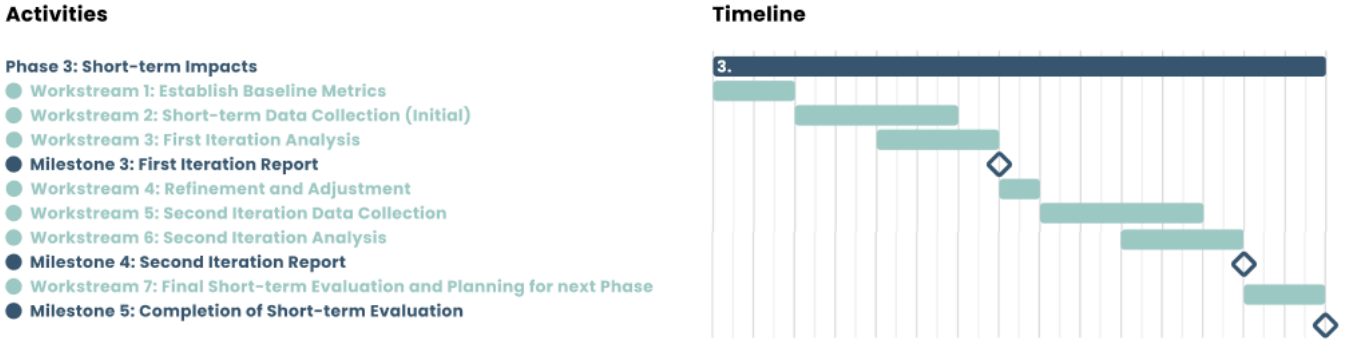


Figure 5: Phase 4 proposed quasi-experimental impact evaluation plan Gantt Chart

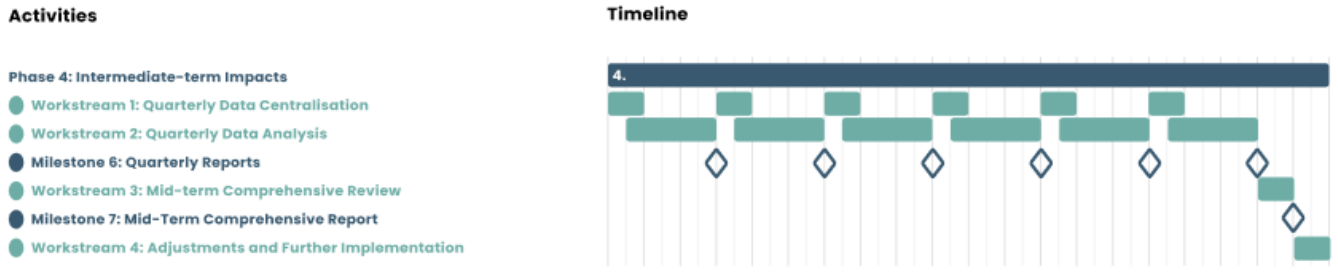
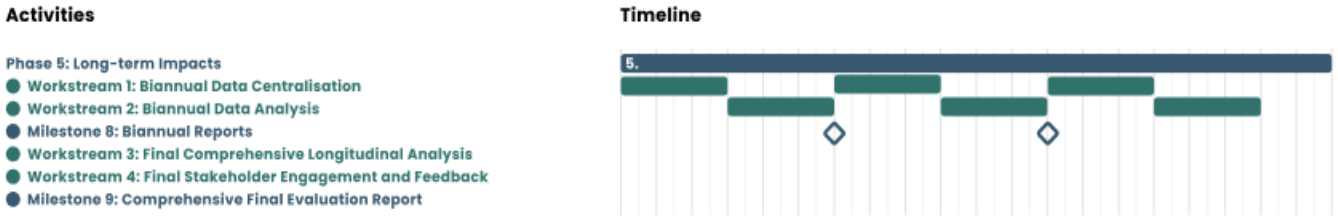


Figure 6: Phase 5 proposed quasi-experimental impact evaluation plan Gantt Chart



Section 4 Value-for-Money evaluation outline

The Value for Money (VfM) evaluation is a critical component of the SCA evaluation process, aimed at determining the economic, efficiency, and ethical justification for the SCA programme. This section presents a series of methodologies designed to provide a comprehensive analysis of the programme's performance across various dimensions.

The National Audit Office (NAO) offers the 4Es framework - Economy, Efficiency, Effectiveness and Equity - to assess the value for money of public sector projects. By applying this framework, the SCA can be assessed on its ability to minimise costs (Economy), maximise outputs from given inputs (Efficiency), achieve intended outcomes (Effectiveness), and ensure fair and accessible services for all users (Equity). This comprehensive evaluation will help ensure that the SCA not only meets its financial goals but also aligns with broader public service objectives, maintaining a balance between cost-saving and quality of service.

Applying this framework when conducting this stage of the evaluation allows for the VfM assessment to go beyond traditional cost analysis and encompass a broader perspective on the qualitative and quantitative benefits of the SCA, ensuring that the investments are yielding proportional and justifiable returns.

Purpose

The primary purpose of the VfM evaluation is to ensure that the SCA team can:

- demonstrate accountability, as providing clear accountability for the use of public funds and resources ensures that these are used efficiently and effectively
- optimise resource use, as identifying opportunities to improve the economy and efficiency in the deployment of the SCA's resources is a key learning from any evaluation and can be applied to the continuous improvement of the SCA's operations as well as other comparable programmes
- validate programme objectives, as this allows the evaluation to confirm that the programme's objectives are being met effectively and that the outcomes align with the strategic goals of HMRC

Benefit types

Cost reduction

The channel shift from traditional customer interaction channels to digital platforms will reduce the need for physical infrastructure and personnel costs, thereby lowering operational expenses and improving VfM. By reducing dependency on phone and post, the SCA can decrease long-term costs related to handling and storage, contributing to a more efficient use of resources. Digital interactions are typically quicker and can handle multiple customers simultaneously, which increases efficiency and drives down costs per transaction, enhancing the overall VfM for HMRC.

Modernised technology

Utilising modern technology allows the SCA to automate routine processes, which not only speeds up operations but also reduces the likelihood of human error, thus optimising resource allocation. The integration of cutting-edge technologies can future-proof the system, making subsequent upgrades less costly and disruptive, thereby improving the long-term VfM. By simplifying tax processes, the SCA can potentially reduce the need for specialised training and support, lowering ongoing operational costs and improving VfM.

Positive user experience

Enhancing the user interface to make it more intuitive can reduce the number of user errors and subsequent support calls, which directly lowers operational costs and improves VfM. A positive user experience increases user engagement and retention, which is likely to lead to higher compliance and more accurate data collection, thereby enhancing the effectiveness of the SCA. Satisfied users are more likely to use digital services repeatedly, reducing the reliance on more costly service channels and thereby improving VfM.

Compliance efficiency

Increasing automation through the SCA leads to more accurate tax submissions, which reduces the need for subsequent corrections and interventions, thus enhancing VfM. Enhanced compliance efficiency through digital services minimises legal and financial risks associated with non-compliance and errors, which can have substantial cost implications for HMRC. By ensuring more reliable and timely compliance, the SCA can improve revenue collection rates, directly impacting the financial health of the nation and ensuring better VfM.

Improved disbursement

Streamlined disbursement processes via the SCA can lead to quicker payouts and refunds, increasing taxpayer satisfaction and trust in the system, which indirectly boosts the perceived VfM. Efficient disbursement mechanisms reduce administrative overheads and the risk of payment errors, both of which are crucial for maintaining cost-effectiveness and VfM. Automation in disbursement can improve the accuracy and traceability of transactions, reducing fraud and leakage, thereby enhancing VfM.

Staff satisfaction

By reducing mundane and repetitive tasks, the SCA can improve staff satisfaction by enabling employees to focus on more complex and rewarding work, thereby increasing productivity and VfM. Higher staff satisfaction can lead to lower turnover rates, reducing recruitment and training costs, which are significant contributors to overall VfM. Satisfied staff are more likely to provide better service to taxpayers, improving user satisfaction and compliance, both key metrics in the VfM assessment.

Environmental sustainability

Reducing paper consumption through digital processes not only aligns with environmental sustainability

goals but also reduces costs associated with printing, storing, and disposing of paper, thereby improving VfM. Digital documentation facilitated by the SCA can be more easily searched and managed, reducing time and resource expenditure and enhancing operational efficiency. The move to digital platforms can help HMRC meet government-wide sustainability targets, improving the organisation's public image and long-term VfM.

Software savings

Consolidating multiple systems into a single digital account can reduce software licensing fees, maintenance costs, and the need for multiple vendor contracts, directly impacting VfM. By using shared platforms and resources, the SCA can reduce the per-user cost of its digital services, enhancing VfM. The scalability of modern software solutions means that increasing user numbers does not necessarily increase costs at the same rate, improving VfM as the user base grows.

Potential unintended consequences

Increased cross-functional collaboration

The need to integrate multiple tax systems into a single platform may foster greater collaboration between different departments within HMRC. This could break down silos, leading to more cohesive and unified approaches to service delivery and problem-solving.

Boost to digital literacy

The push towards using the SCA may encourage users to improve their digital skills. For many, particularly older or less technologically adept individuals, this interaction might catalyse a broader engagement with digital services, improving their overall digital literacy.

Unintended regulatory compliance

The design of the SCA might inadvertently align with other compliance needs or upcoming regulations, providing HMRC with a head start in meeting new regulatory requirements without additional adjustments.

Scalability for future needs

The architecture of the SCA, initially aimed at simplifying user access and system integration, could unexpectedly prove ideal for scaling up or adapting to future changes in tax law or policy without significant reconfigurations.

Rapid and seamless implementation of changes

With a centralised and streamlined system like the SCA, HMRC can implement changes in tax policy, rates, or procedures quickly and uniformly across all interfaces. The benefit here is the minimal disruption to the user, who may not need to adjust or learn new processes to comply with these changes, thus reducing confusion and ensuring smoother transitions.

Cost types

Direct costs

The following direct costs have been identified:

- development and implementation costs, including expenses related to software development, integration, and deployment of the SCA system, as well as hardware, software licences, and the labour costs of IT professionals and consultants
- training costs, including the time and resources required for training employees across HMRC on the new system in order to ensure a smooth transition and effective use
- maintenance and operational costs, which includes ongoing costs to maintain and update the system, manage data security, and provide user support
- transition costs, including the downtime or reduced productivity resulting from staff adapting to systems as they switch from old platforms to the SCA
- compliance and regulatory costs, including additional consulting and legal services, especially in areas like data protection to ensure the SCA meets all legal and compliance requirements

Attribution of SCA's effects

Accurate attribution of benefits and costs through a rigorous impact evaluation is crucial for conducting a comprehensive VfM assessment, especially in public sector initiatives like the SCA. In the context of VfM, rigorous impact evaluations extend their relevance beyond financial metrics to include broader outcomes, enhancing public accountability. For HMRC, capturing the full scope of the SCA's impact informs necessary adjustments and validates the project's contribution to public service. This detailed understanding supports transparency and aids in decision-making for future projects, ensuring resources are utilised effectively to achieve the maximum possible public benefit. Overall, such evaluations are vital in demonstrating that the investment in the SCA is prudent and beneficial from multiple perspectives.

Section 5 Implementation and timescales

Details for the execution of each evaluation type can be found at the end of the corresponding outline, however, in this section, we cover wider considerations to be made for the implementation of the full-scale evaluation. This includes the skills and capabilities required and further recommendations for the procurement of the evaluation if conducted by an independent evaluation team.

Skills and capabilities

Whether conducted internally or by an external organisation, several skills and capabilities are required within the evaluation team to execute the delivery plan. These skills have been divided by the recommended project role below.

Project and delivery management

This includes overseeing major research projects, managing multidisciplinary teams, risk and timeline management, external stakeholder engagement and problem-solving.

Statistical and data analysis

This includes using different data types (e.g. geospatial data) and impact evaluation approaches, conducting statistical matching, and using statistics to derive causality.

Qualitative and social research methods

This includes designing surveys and using data collection tools, conducting interviews, focus groups and workshops for user research, and coding qualitative responses.

Economic analysis

This includes conducting CBA and SCEA approaches, conducting NPV assessments, cost accounting, and evaluating the potential risks associated with economic decisions.

Technical advisory

This includes sufficient technical expertise for appraising technical development and architecture decisions, and assessing associated value for money.

Procurement

In addition to the time taken to prepare internal specifications and approaches, the SCA team should factor in procurement timeframes if they decide to appoint an independent evaluation team, either to deliver or support the evaluation. For an open tender, timeframes for conducting a procurement competition typically last between one and three months, with the time taken to conduct selection, award, standstill and contract onboarding processes.

The use of a framework, such as Crown Commercial Service's (CCS)' Research and Insights Dynamic Purchasing System (DPS), can expedite procurement timeframes and has been a common route for commissioning independent evaluations. The majority of procured monitoring and evaluation projects are done so through the Research and Insights DPS (RM6126). It is worth noting that most monitoring and evaluation projects are typically procured as separate process, impact and VfM evaluations.

Annex A: Process evaluation methods

In this section, we summarise the individual methods of data collection mentioned above. For each method, we provide a stepwise guide for executing the process evaluation.

We recommend that the future process evaluators of the SCA programme employ a mixed-methods approach, incorporating each of the methods described in this section were identified as most appropriate in the process evaluation outline. Additionally, we understand that the HMRC already conducts user research alongside SCA development and delivery, and we therefore encourage future evaluators to align process evaluation efforts with existing user research engagements.

Further best practices for the collection of data via all methods mentioned in this process evaluation outline can be found in Chapter 4: Data collection, data access and data linking of The Magenta Book.

In-depth interviews

In-depth interviews gather detailed data through one-on-one conversations. They can provide insights into interviewees' experiences, attitudes and perceptions. Interviews hold a comparative advantage in their ability to explore topics deeply, allowing for a nuanced understanding of complex phenomena, and surpassing the surface-level data obtained through other methods like surveys or focus groups. Interviews are the preferred method of collecting feedback, as ensured anonymity allows for participants to speak freely. This form of data collection can be performed with the following steps.

Step 1: Define process evaluation research questions

This involves:

- identifying target population
- developing a semi-structured interview guide (include questions to be asked of all interviews but allocate time for open-ended questions) Develop codebook
- identifying preferred coding software
- developing consent forms

Step 2: Select participants

This involves:

- scheduling interviews
- determining variation across relevant participant characteristics (e.g. level of seniority of programme officers, and function of team members)
- obtaining consent

Step 3: Conduct interviews

This step involves:

- Recording interviews (if online, record on a video platform, if in person, record on other suitable implement)
- Revising the codebook if needed

Step 4: Analyse data

Surveys and questionnaires

Surveys and questionnaires are used when collecting data from a large number of respondents, where the goal is typically to obtain information that is representative of the general population. These generate quantitative data through standardised response options and have the benefit of providing respondents with anonymity. This form of data collection can be performed with the following steps.

When conducting surveys a sampling method must be chosen, so that results can be systematically framed. For responses that are being collected from the Delivery or Development teams themselves, we suggest Stratified Random Sampling where the strata are different teams or subunits within the broader Delivery and Development functions. For responses that are being gathered from SCA's end-users, we suggest random sampling to ensure that the responses are representative of the total population.

The first step is to use process evaluation questions to define research objectives. The next step is to design the survey. This involves developing salient questions, choosing an answer format (for example, multiple choice, Likert Scale, open-ended), piloting the survey with test groups to identify issues with comprehension, clarity and relevance of questions and their wording. The third step is to select a sampling method, after determining a target population to whom the results can be generalised. Then, the survey can be administered. This involves matching the disbursement method (for example, online, telephone, in-person) to the target population, and designing clear instructions for completing the survey, matched to the target population. The final steps involve cleaning and preparing the data, as well as analysing it.

Focus groups

Focus groups collect data from groups of participants who engage in a structured conversation around a central topic, guided by a moderator. This method allows for the recovery of consensus or identification of conflict lines around a topic and is best suited for answering solutions-oriented questions. These can be carried out with the following steps.

The first step is to define process evaluation questions, determine the ideal focus group composition and develop a discussion guide. The next step is to select participants, schedule interviews and determine variation across relevant participant characteristics (e.g. level of seniority of programme officers, and function of team members). Conducting the focus group involves recording interviews. The final step is to analyse the data, by obtaining focus group transcripts and identifying recurrent themes and patterns of responses.

Programme documentation review

Programme documentation review as a research method involves systematically recording and analysing information about an output. It includes various documents such as programme manuals, reports,

policies, and other written materials. The process typically involves collecting, organising, and analysing these documents to understand the programme's goals, implementation strategies, and output performance data. A programme documentation review can be conducted with the following steps.

The first step is to choose a process evaluation research question and then identify relevant programme documentation by conducting interviews with the programme team. The final step involves organising data, by tracking which document pertains to which research question/aspect of the research question and identifying data gaps as they occur, before moving on to analysing this data.

Annex B: Impact evaluation methods

In this section, we summarise the practical steps to take for each of the methods described in the impact evaluation outline of this evaluation plan, including, instrumental variables, and propensity score matching.

Conducting instrumental variables analysis

To conduct an instrumental variable analysis, we outline the following steps:

Preliminary Data Analysis

This involves compilation and preliminary analysis of data to understand patterns of SCA usage and potential influences.

Actions include collecting historical data on SCA usage and related variables, as well as conducting statistical analyses to identify trends and correlations.

The output is a report detailing the data landscape and identifying potential variables for instrument testing.

Instrument Testing and Validation

This involves rigorous testing of the proposed instruments to ensure they meet the necessary statistical criteria for a valid instrument.

Actions include implementing correlation tests to assess the relationship between potential instruments and SCA usage, and conducting exclusion restriction tests to confirm that these instruments do not directly affect the outcomes.

A validation report confirming the suitability of selected instruments should be produced, supported by statistical evidence and theoretical justification.

Two-stage least squares (2SLS) analysis

Execution of the 2SLS methodology to estimate the causal effects of SCA using the validated instruments.

Actions include applying the first-stage regression to predict SCA usage based on the instrument, followed by a second-stage regression to assess the impact of predicted SCA usage on outcomes.

A detailed analysis report providing estimates of the causal impact of SCA should be produced, including confidence intervals and significance levels.

Results compilation and evaluation

This requires synthesis of analysis results into a comprehensive evaluation report that integrates findings

and offers actionable insights.

Actions include compiling the results from the 2SLS analysis, interpreting the findings in the context of HMRC's objectives, and drafting recommendations based on the study's conclusions.

This should lead to an evaluation report that provides clear insights into the causal impacts of SCA, discusses methodological considerations, and outlines recommendations for policy and programme improvements.

Conducting propensity score matching analysis

To conduct a propensity score matching analysis, we outline the following steps.

Propensity score development

Calculate propensity scores for all individuals in the dataset based on their likelihood of using the SCA, using logistic regression or another appropriate model.

Actions include collecting data on covariates likely to influence SCA usage, modelling the probability of SCA adoption, and computing scores.

This leads to the output of a Propensity Score Calculation Report that documents the model used, the scores calculated, and the distribution of scores across the sample.

Matching and data preparation

Use the propensity scores to match SCA users with non-users or less frequent users who have similar scores.

Key actions include implementing a matching algorithm such as nearest neighbour matching, ensuring that matches are within a predefined range to maintain similarity.

A Matched Dataset that pairs treated and control units based on their propensity scores should be produced, ready for outcome analysis.

Outcome Comparison Analysis

Analyse and compare outcomes between matched pairs to assess the impact of the SCA.

This requires conducting statistical tests to compare outcomes like service usage and compliance between the matched groups.

This should lead to the production of an Impact Analysis Report detailing the differences in outcomes between the treatment and matched control groups, providing evidence of the SCA's effectiveness.