# Evaluation of the UKRI Future Leaders Fellowships

## Evaluation framework report

RAND Europe

November 2022
Prepared for UKRI

# Executive summary

The development and retention of research talent is key to the delivery of a wide range of UK Government strategies and commitments, including the Industrial Strategy, the Plan for Growth, the Life Sciences Vision, and the UK Innovation Strategy. The UK Research and Development (R&D) Roadmap and the R&D People and Culture Strategy articulate the UK's goals surrounding support for researcher careers and capabilities and the positioning of the UK as a world leader in attracting and retaining the best talent. UKRI, as the UK's largest funder of R&D, is central to achieving the outcomes set out in these strategies, with the Future Leaders Fellowship (FLF) programme as a key mechanism for delivery on these goals. The FLF, a £900 million investment, occupies a unique position within the landscape of early-career researcher programmes in the UK, supporting both UK and international researchers working in a wide range of disciplines and sectors, and encouraging novel and interdisciplinary research projects.

In June 2022, UKRI commissioned RAND Europe to conduct an evaluation of the FLF programme. The aims of this evaluation are to understand the impact of FLF, draw lessons regarding how programme implementation has enabled impact, and explore the programme's value for money. This report (the first project deliverable) presents the evaluation framework for the evaluation of the FLF. The report has been prepared to guide the evaluation and for UKRI to approve the planned approach prior to commencement of the core evaluation data collection and analysis tasks.

The proposed evaluation approach is underpinned by two primary analytical frames: the FLF Theory of Change and an adapted version of the Kirkpatrick model, a framework traditionally used for evaluating training interventions. Our evaluation approach will comprise: a process evaluation; an impact evaluation comprising a combination of quasi-experimental design and theory-based evaluation; and an economic evaluation (subject to further feasibility assessment).

The process evaluation will aid understanding of how programme design, delivery and support has been received, whether the programme targets have been met and identify key challenges or barriers, if any, thereby identifying opportunities for learning. It will also assess whether the FLF has been designed and implemented in a way that supports the anticipated impacts of the programme.

The quasi-experimental impact evaluation will aid in assessing the extent to which selective outcomes have been achieved by FLF fellows compared to a proposed counterfactual group. This comparative assessment will be supplemented by a crucial theory-based evaluation which will capture more qualitative outputs and outcomes and explore whether the FLF is on track to achieve sustained longer-term impact. The theory-based component will include a realist analysis of the context and the mechanisms through which a given set of outcomes have been achieved for the fellows. This approach has the advantage of not only providing evidence of whether FLF has generated impact but also of how it has done so and for whom, so that the programme and others like it can be further improved in the future.

An economic feasibility assessment, to be undertaken in parallel with the impact evaluation, will focus on assessing the viability of undertaking a cost benefit analysis and exploring mechanisms for assessing value and effectiveness.

# Table of contents

# Abbreviations

AHSS      Arts, humanities and social sciences

BEIS      Department for Business, Energy and Industrial Strategy

BSD       Business Structure Database

CMO       Context Mechanism Outcome

EDI       Equality, diversity and inclusion

FLF       Future Leaders Fellowships

G7        The Group of Seven

GIAA      Government Internal Audit Agency

HE-BCI    Higher Education Business and Community Interaction

HESA      Higher Education Statistics Agency

IES       Institute for Employment Studies

MIDRI     Multidisciplinary and interdisciplinary research and innovation

MRC       Medical Research Council

NPIF      National Productivity Investment Fund

ONS       Office for National Statistics

QED       Quasi-experimental design

R&D       Research and Development

R&I       Research and Innovation

STEM      Science, technology, engineering and mathematics

TBE       Theory-based Evaluation

ToC       Theory of Change

UK        United Kingdom

UKRI      UK Research and Innovation

# 1. Introduction

This report sets out the evaluation framework that will be used to guide the evaluation of the FLF. The report is structured as follows:

- The remainder of this Introduction describes the context for the FLF, the goals and structure of the programme, the aims of the evaluation and our approach to developing this evaluation framework report.

- Chapter 2 presents the analytical framing for the evaluation comprising the FLF ToC (in both full and simplified forms) and the Kirkpatrick model.

- Chapter 3 presents our overarching evaluation approach through which we will collect evidence to assess the FLF against the analytical frameworks. The evaluation approach will comprise three types of evaluation: a process evaluation; an impact evaluation comprising both a quasi-experimental design (QED) and a theory-based evaluation (TBE); and an economic evaluation.

- Chapter 4 sets out considerations for the QED evaluation based on the evaluation team's further scoping during the evaluation framework stage.

- Chapter 5 presents the process and impact evaluation frameworks through which the evaluation of the FLF will be operationalised. Presented in tabular form, each framework delineates evaluation questions (EQs) derived from the FLF ToC which our evaluation will seek to answer, together with metrics and proposed data sources for each EQ.

- Chapter 6 presents the plan and timescales for the evaluation, including the dates of all key evaluation deliverables.

- Chapter 7 outlines key limitations and risks of our approach to the evaluation of the FLF.

- The Annexes to the report contain: additional information on the document review undertaken for this report (Annex A); a more detailed version of the impact evaluation framework (Annex B); and further information on the QED counterfactual generation (Annex C).

## 1.1. Development and retention of research talent in the UK

The UK punches above its weight as a world leader in R&I despite comprising only one percent of the world's population. According to UKRI, the UK accounts for seven percent of the world's academic

publications and 14 percent of the world's most highly cited academic publications.[1] The UK has also held the title for the highest field-weighted citation impact amongst all the G7 countries since 2007,[2] and is consistently ranked in the top five countries for innovation.[3]

Talent development and retention is a key part of the UK's status as a world leader in R&I. The UK's performance is underpinned by its ability to support researcher careers and capabilities and its positioning as a world leader in attracting and retaining the best talent. The need to continue developing, attracting, and retaining research talent has been articulated in a wide range of UK Government strategies, including the UK Industrial Strategy,[4] the Plan for Growth,[5] the Life Sciences Vision,[6] the UK Innovation Strategy,[7] and Sector Deals promoting government and industry partnerships.[8]

In recent years, the UK Government has introduced new policies and investments aimed at making the country a more attractive destination for R&I. The UK Research and Development (R&D) Roadmap, published in 2020, highlights the importance of attracting and retaining diverse talent in the country.[9] The R&D Roadmap touches upon a number of key areas for improvement, including levels of public and private investments, bureaucratic processes, work culture and careers, development and innovation, regional imbalances, and the international context.[10] As such, the UK Government has published a number of R&D strategies aimed at better supporting research talent and has committed to increase investment in R&D to 2.4% of GDP by 2027 and increase public funding for R&D to £22 billion per year by 2024/2025.

In its 2021 R&D People and Culture Strategy, the UK Government also sets out a comprehensive vision for how the UK will develop, attract, and retain research and innovation talent to build on existing strengths and meet future challenges.[11] The Strategy outlines three main pillars: people, culture, and talent.

- Regarding 'people', the Strategy estimates that the R&D sector will need at least an additional 150,000 researchers and technicians by 2030 to sustain UK's target of 2.4% R&D intensity. Challenges identified include narrow career paths and variable quality of leadership. To address this, the Government is developing a New Deal for post-graduate students and the Young Academy

[1] UKRI, "Science & Research," n.d., https://greattalent.campaign.gov.uk/work-in-the-uk/science-research/.

2 UKRI.

3 UKRI, "Science & Research."

[4] HM Government, "Industrial Strategy: Building a Britain Fit for the Future," 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf.

[5] HM Treasury, "Build Back Better: Our Plan for Growth," March 3, 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/968403/PfG_Final_Web_Accessible_Version.pdf.

[6] HM Treasury, "Life Sciences Vision," July 6, 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1013597/life-sciences-vision-2021.pdf.

[7] BEIS, "UK Innovation Strategy: Leading the Future by Creating It," July 22, 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1009577/uk-innovation-strategy.pdf.

[8] BEIS, "Introduction to Sector Deals," 2019, https://www.gov.uk/government/publications/industrial-strategy-sector-deals/introduction-to-sector-deals.

[9] HM Government, "UK Research and Development Roadmap," July 1, 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896799/UK_Research_and_Development_Roadmap.pdf.

[10] HM Government.

[11] BEIS, "R&D People and Culture Strategy: People at the Heart of R&D," July 22, 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004685/r_d-people-culture-strategy.pdf.

programmes, and is also encouraging more people into R&I careers by providing support for flexible, cross-sector training programmes, supporting interdisciplinary approaches, evaluating expert peer review approaches, evaluating the impact of research grants, and promoting the development of leadership and management skills.

- Regarding 'culture', the Strategy identifies challenges including a lack of inclusivity, incentives, and public collaboration, along with unnecessary bureaucracy. In response, the Strategy seeks to promote positive, inclusive and respectful cultures, recognise and reward people and activities that lead to excellent R&I, tackle issues of bullying and harassment, and increase public engagement.

- Regarding 'talent', the Strategy aims to improve and promote funding offers to attract and support the best researchers, work with funders and the R&I community to open up opportunities, and facilitate migration routes for innovators, entrepreneurs, and top talent. To this end, the new cross-department Office for Talent is being established under the 2020 Roadmap to offer support for and address barriers to attracting and retaining global innovation talent to the UK, provide key information and make it easier for talent to come to the UK.

## 1.2.    The UKRI Future Leaders Fellowships as a key investment

As the coordinating body for UK R&I funding, UKRI is central to achieving the outcomes set out in policies such as the UK R&D Roadmap and the People and Culture Strategy. Activities undertaken by UKRI to support these outcomes are wide-ranging, including leading a cross-sectoral consultation to inform development of a New Deal for post-graduate research students, launching initiatives to promote equality, diversity and inclusion (EDI) in research, and reinvigorating participation in the peer review system.[12] UKRI is also a signatory to the Researcher Development Concordat and has committed to driving change in the culture of R&I.[13]

Alongside such activities, the Future Leaders Fellowships (FLF) programme represents a key investment for the delivery of the UK's researcher development goals. Announced in 2018, FLF is a £900 million fund that seeks to support the careers of world-class researchers across UK business and academia. Underpinned by funding from the National Productivity Investment Fund (NPIF), the FLF programme aims to establish the careers of R&I leaders across business and academia, improve UK R&I, foster and retain talent in the UK, and develop new and better career pathways. The specific aims and objectives of the FLF are to promote:

- High quality and impactful research and innovation in areas aligned with the Government's Industrial Strategy

---

[12] UKRI, "New Deal for Postgraduate Research," June 24, 2022, https://www.ukri.org/what-we-offer/developing-people-and-skills/new-deal-for-postgraduate-research/.

[13] UKRI, "Concordat to Support the Career Development of Researchers: UKRI Funder Action Plan," July 2020, https://www.ukri.org/wp-content/uploads/2020/10/UKRI-071020-ConcordatToSupportTheCareerDevelopmentOfResearchersFunderActionPlan.pdf.

- Increased engagement between industry and academia on research and innovation activities, including through collaboration and problem-solving, and facilitating the movement of people between sectors

- Increased multi- and interdisciplinary research and innovation

- Develop, retain, attract and sustain highly skilled research and innovation leaders of the future, from within the UK and from overseas

- A more equal, diverse and inclusive research and innovation workforce, which welcomes international talent

- Provide sustained funding and resources for the best early career researchers and innovators to tackle difficult and novel challenges and deliver value for money.

While existing alongside a number of other early career fellowship schemes, such as the Leverhulme and British Academy schemes, FLF occupies a unique position within the R&I landscape. Most other schemes focus on specific disciplinary or subject area domains. FLF, by contrast, is a pan-UKRI programme supporting fellows working in a wide range of disciplines and sectors. Under FLF, UKRI has developed new approaches to allow more fellows to work with and in businesses (as well as public sector and charitable bodies), while also encouraging novel and interdisciplinary projects. Moreover, unlike most other schemes, FLF has emphasised long-term support to researchers – offering four to seven years of funding – with the aim of facilitating the transition of talented researchers into positions of leadership and independence. With 40% of awards going to non-UK nationals, FLF is also helping attract global talent to the UK in ways not done through traditional fellowship schemes.

## 1.3.   The FLF portfolio

FLF funding will be allocated over an 11-year timeframe between 2018-2029. Between 2019 and 2022, six rounds of funding have been awarded, with 499 fellowships awarded as of March 2022. Each FLF award is for an initial period of four years, with an option for an additional three years of funding.

**Figure 1: Timeline of FLF rounds (dark grey represents initial funding period and light grey the optional three-year extension period of each round of awards)**

| | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FLF - Round 1 | | | | | | | | | | | | |
| FLF - Round 2 | | | | | | | | | | | | |
| FLF - Round 3 | | | | | | | | | | | | |
| FLF - Round 4 | | | | | | | | | | | | |
| FLF - Round 5 | | | | | | | | | | | | |
| FLF - Round 6 | | | | | | | | | | | | |

Source: UKRI

In funding rounds 1-6 of FLF, most fellowships were awarded to researchers within academic institutions (Figure 2). However, the number of business fellows has increased year on year, reaching a high of 13 in Round 6 (16 percent of awards within the funding round).

**Figure 2: FLF awards by sector and funding round**
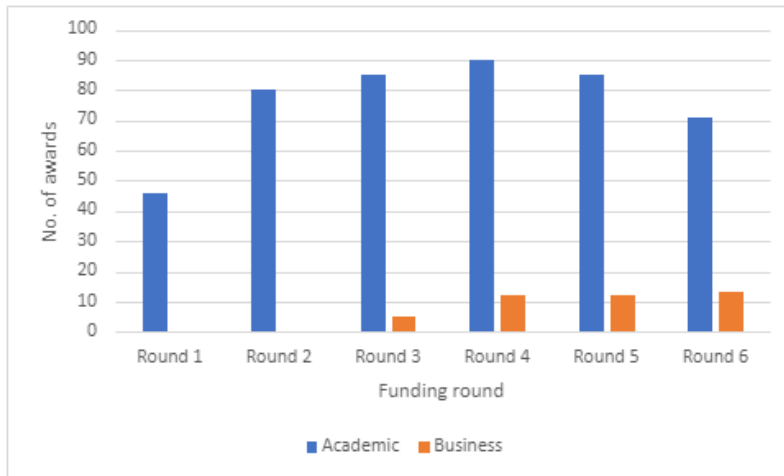


Source: RAND Europe analysis of UKRI data

The FLF portfolio covers a wide range of disciplines and subject areas. Figure 3 shows the primary Research Council (RC) for FLF fellowships awarded in funding rounds 1-6. This illustrates that while EPSRC and MRC have been the primary RC for the highest number of awards, each funding round has covered a wide breadth of RCs, including both science, technology, engineering and mathematics (STEM) and arts, humanities and social sciences (AHSS) focused councils.

**Figure 3: FLF awards by primary Research Council and funding round**
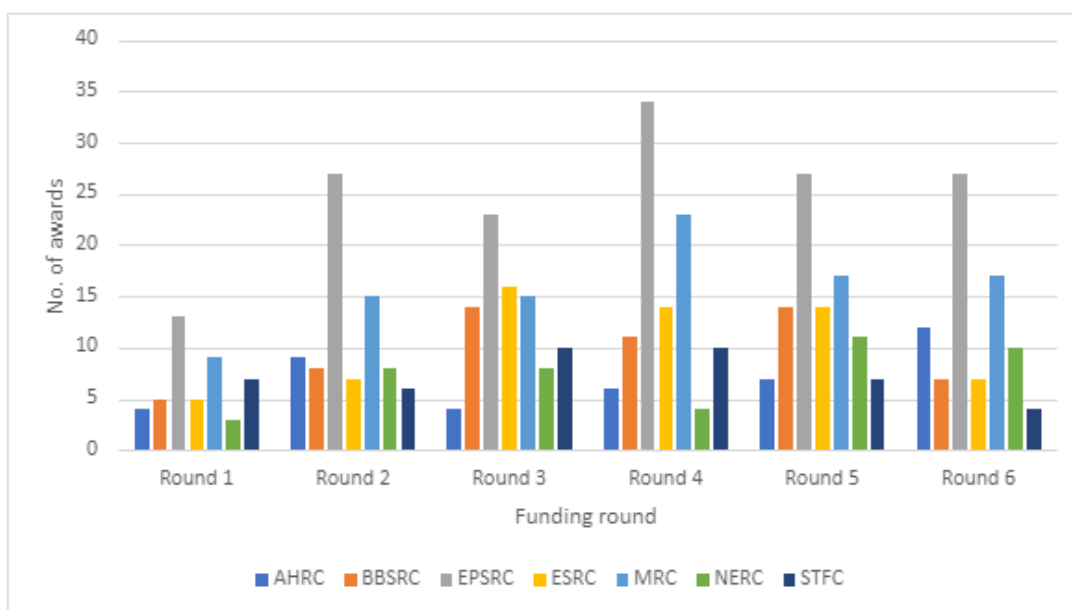


Source: RAND Europe analysis of UKRI data

Figure 4 shows the top 10 subject areas[14] in which most ISCF awards have been funded overall across funding rounds 1-6.[15] While the highest number of awards have been to fellows working within Biochemistry and Cell Biology, the figure also illustrates that FLF fellowships have covered a diverse mix of subject areas. Notably, STEM subjects are predominant within the top 10 funded subject areas.

**Figure 4: FLF awards by subject area (10 most funded subject areas[16]) and funding round**
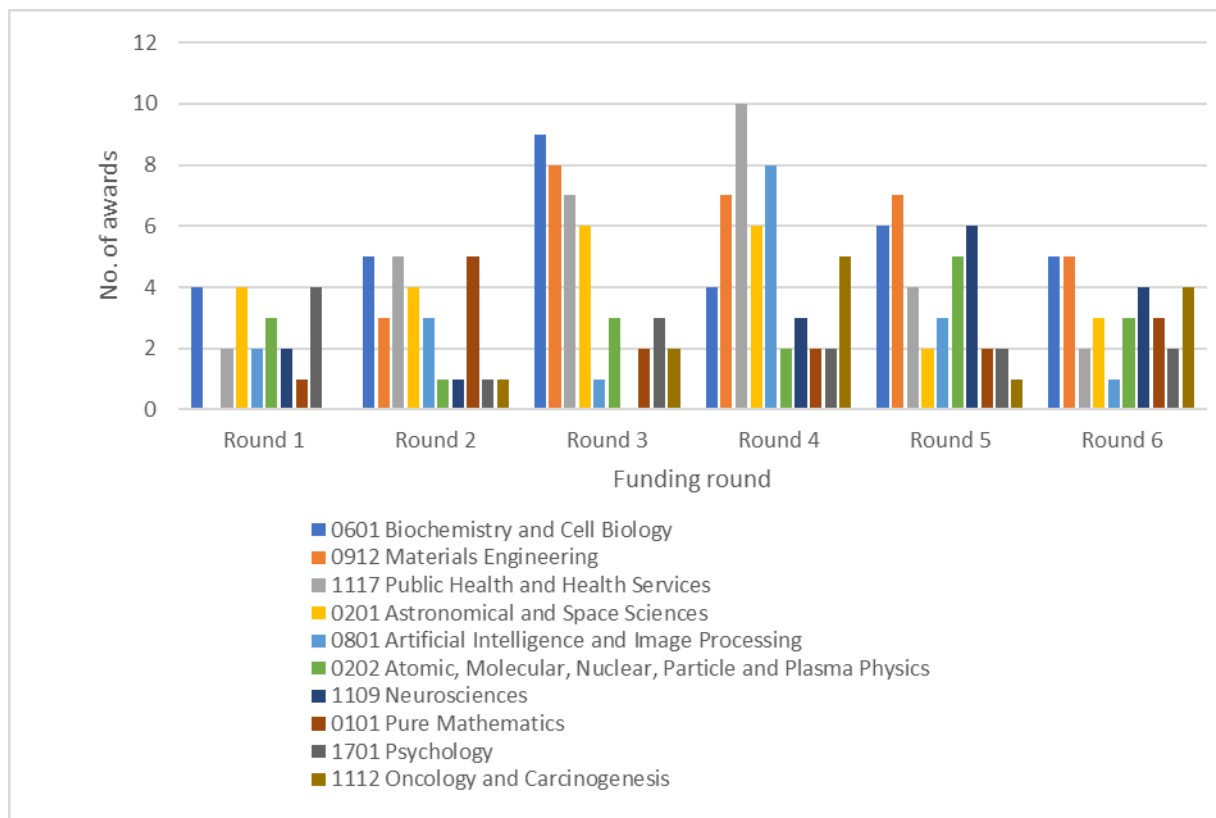


Source: RAND Europe analysis of UKRI data

## 1.4. Evaluation of the FLF

UKRI is committed to rigorous evaluation of the FLF. In 2020, following a UKRI workshop conducted to explore the FLF objectives and identify key challenges for the evaluation of the programme, UKRI commissioned a scoping and feasibility study on the evaluation of the FLF. Conducted by the Institute for Employment Studies (IES), the scoping and feasibility study considered appropriate methods and approaches for answering the evaluation questions identified by the UKRI workshop.[17] The key objectives of the study were to:

---

[14] Subject areas are those used by the Dimensions database.

[15] The figure shows the 10 subject areas in which the highest number of fellowships have been awarded overall across the first six funding rounds. These top 10 overall subject areas are then broken down by funding round. The number of awards outside these top 10 subject areas are as follows: Round 1 – 24 awards; Round 2 – 51 awards; Round 3 - 49 awards; Round 4 – 53 awards; Round 5 – 59 awards; Round 6 – 52 awards.

[16] Subject areas (and numerical references) used are those of the Dimensions database (https://www.dimensions.ai/).

[17] Emma Pollard et al., "UKRI Future Leaders Fellowships Evaluation Scoping & Feasibility Study" (Institute for Employment Studies, April 30, 2021), https://www.ukri.org/wp-content/uploads/2022/05/UKRI-12052022-FLF-SF-FINAL-amended-230222.pdf.

- develop a summary of existing knowledge and information about potential evaluation methods

- explore challenges around evaluating the FLF programme (including the trade-offs between evaluation approaches)

- make recommendations for a full evaluation of the FLF.

Following the scoping and feasibility report, in June 2022, UKRI commissioned RAND Europe to conduct a full evaluation of the FLF. The aims of this full evaluation are to understand:

- whether and how the FLF scheme delivered on its intended aims and objectives

- to what extent the programme achieved its planned impacts, as outlined in the ToC

- for whom and in what circumstances did it achieve this

- what factors in the context or operation of FLF supported or inhibited the emergence of impacts

- whether the intervention is cost-effective and a justifiable public investment.

The evaluation will draw upon the insights and recommendations of the scoping and feasibility study and comprise a process evaluation and an impact evaluation of the FLF. The evaluation will also further explore the feasibility of an economic evaluation of the FLF. The evaluation will cover rounds 1-6 of the FLF programme. The timeline for the evaluation will be 2022-2024, with a possible extension to 2025 to conduct an economic evaluation, depending on the feasibility of this aspect.

## 1.5.    FLF evaluation framework report

As the first deliverable of the evaluation of the FLF, this document presents the evaluation framework report for the evaluation of the FLF. The purpose of the report is to outline a detailed evaluation approach, including evaluation questions, data collection and analysis methods to provide the basis for robust process and impact evaluation of the FLF. The report has been prepared to guide the evaluation and for UKRI to approve the planned approach prior to the commencement of core evaluation data collection and analysis tasks. Below, we outline the key activities undertaken by the evaluation team that have informed the development of this report.

- **Review of the Institute for Employment Studies (IES) scoping report**

The scoping and feasibility study[18] conducted by IES was reviewed in detail to assess and develop the links between the FLF ToC, the key evaluation questions and themes, and the potential metrics and data sources that can be collected as evidence. Based on this review, we have developed a refined list of evaluation questions, metrics and data sources for the evaluation (the refined evaluation framework is presented in Chapter 5). We have also developed a simplified version of the FLF ToC to draw out more clearly the distinct pathways specified within it, i.e. the host, the fellow, the idea, and diversity and porosity (this simplified version of the ToC is presented in Chapter 3).

---

[18] Emma Pollard et al.

- **Review of management information**

The analysis of management information such as FLF candidate demographics, sift scores and CVs and details of the six rounds of funding was undertaken to consider the options for a QED evaluation (the resulting QED considerations are presented in Chapter 4). The management information was also used to assess the type of metrics that could inform the evaluation. We anticipate using the information further to create bespoke impact data collection tools such as career tracker surveys and key informant interview topic guides. Further review and access to new management information (e.g. Researchfish returns) will be required at subsequent stages to enable refinement of the evaluation methods.

- **Review of existing FLF audits, reporting and reviews**

A review of three types of FLF documents (i) the Government Internal Audit Agency (GIAA) audit; (ii) the BEIS critical friend review; and (iii) BEIS quarterly reports, was conducted with the aim of identifying insights and evidence to inform and refine the evaluation framework with specific reference to the process evaluation. The review consisted of four steps that included identifying relevant FLF documents, creating a review and extraction template based on process framework questions, identifying key trends and themes in the documents, and using this overall insight to refine the process framework questions. More information on each of these steps, including key findings of the review, is presented in Annex A.

- **Steering group and FLF team meetings**

The initial kick-off meeting between the evaluation team and the UKRI evaluation steering group provided additional contextual information about the set-up and delivery of the FLF programme. Follow-up meetings allowed further exchange and discussion of datasets, data sources and access to proprietary platforms that could be provided by UKRI. This has informed the feasibility of our evaluation framework. Follow-up discussions on potential metrics also allowed further iteration of the process and impact frameworks.

# 2. Analytical framing

This chapter sets out the analytical framing for the evaluation of the FLF. Our evaluation design is underpinned by two primary analytical frames: the FLF ToC and an adapted version of the Kirkpatrick model. While the FLF ToC provides an underpinning framework for the assessment of the programme against envisioned inputs, activities, outputs, outcomes and impacts; the Kirkpatrick model will provide an overarching conceptual and analytical lens allowing multifaceted evidence from across the three-part evaluation to be triangulated into a holistic overview.

## 2.1. FLF ToC

An FLF ToC has been developed by IES drawing on a logic model developed by UKRI. The ToC is intended to serve two broad purposes:

- The ToC articulates a shared understanding of the aims of the FLF programme and how these will be achieved, thereby helping support the alignment of FLF stakeholders and beneficiaries.

- For the purposes of the evaluation of the FLF, the ToC also provides an important resource by setting out the envisioned intervention logic against which the performance of the programme can be assessed. In Chapter 5 below, we present an evaluation framework that maps evaluation questions, metrics and data sources to the various elements of the FLF ToC.

While the FLF ToC adopts the typical logic model approach of delineating anticipated inputs, activities, outputs, outcomes and impacts of the programme, it also comprises four intersecting causal pathways: the fellow, the host, the idea, and diversity and porosity. The causal pathways reflect the fact that FLF's impacts are expected to be achieved by multiple 'agents for change' while also capturing the intended 'ripple effects' of the programme, including the creation of more porous, diverse collaborations at all levels. A visual representation of the FLF ToC as developed by IES is presented in Figure 5 below. A fuller narrative explanation of the ToC can also be found within the IES scoping and feasibility study.[19]

---

[19] Emma Pollard et al.

**Figure 5: FLF ToC diagram**
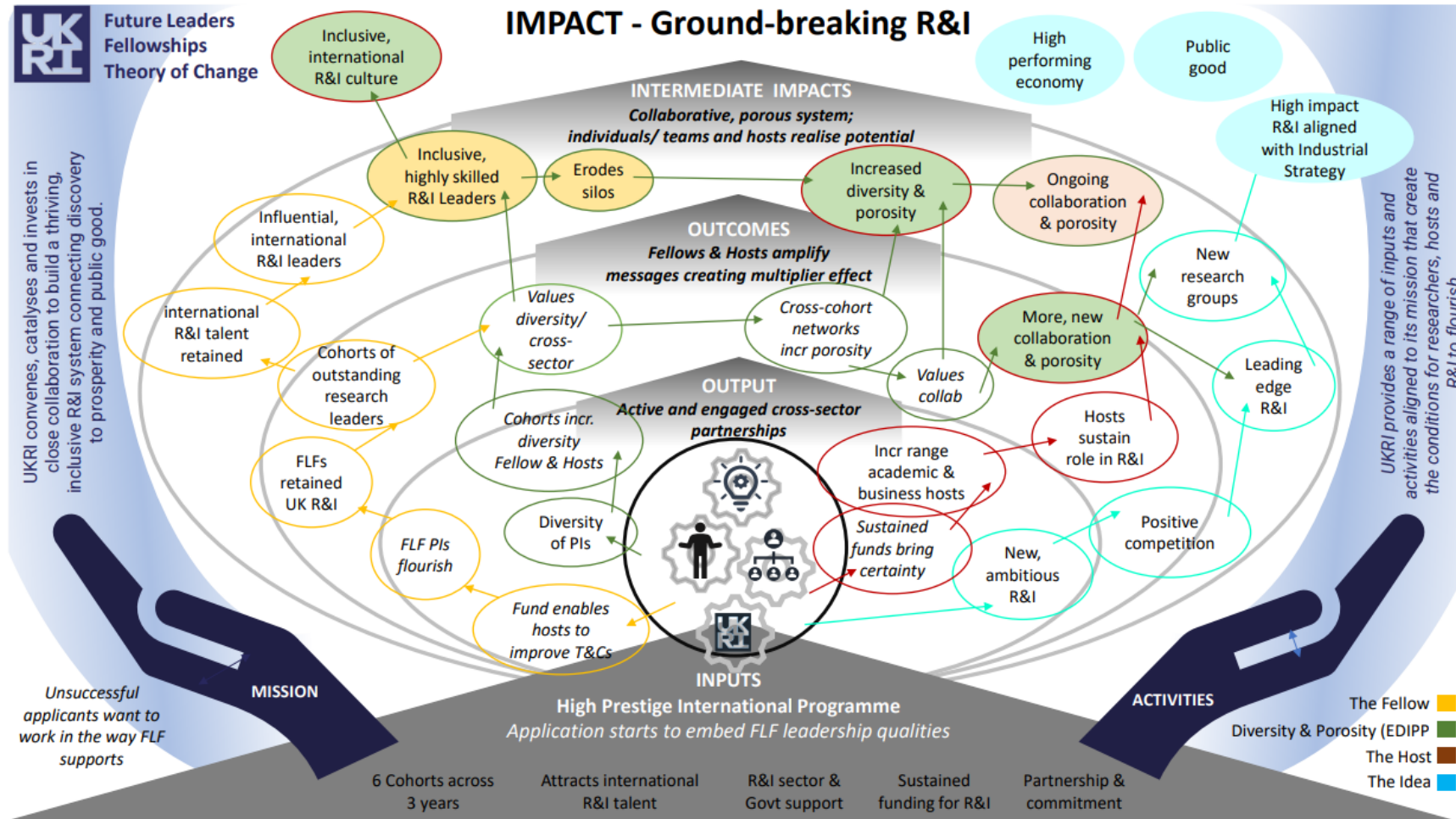


Source: IES scoping and feasibility study

While the FLF ToC presented above provides helpful articulation of the varied and intersecting mechanisms through which the FLF programme is intended to achieve impacts, it is also complex and difficult to digest. In Figure 6 below, we present a simplified version of the FLF ToC. The modified ToC retains the same logic model approach with intersecting causal pathways but synthesises the elements at each level to capture the intended logic of the FLF more succinctly. This simplified ToC has been developed primarily to support the evaluation team in developing an evaluation framework that can appropriately capture the intended impacts of the FLF (as presented in Chapter 5) below.

**Figure 6: Simplified FLF ToC**
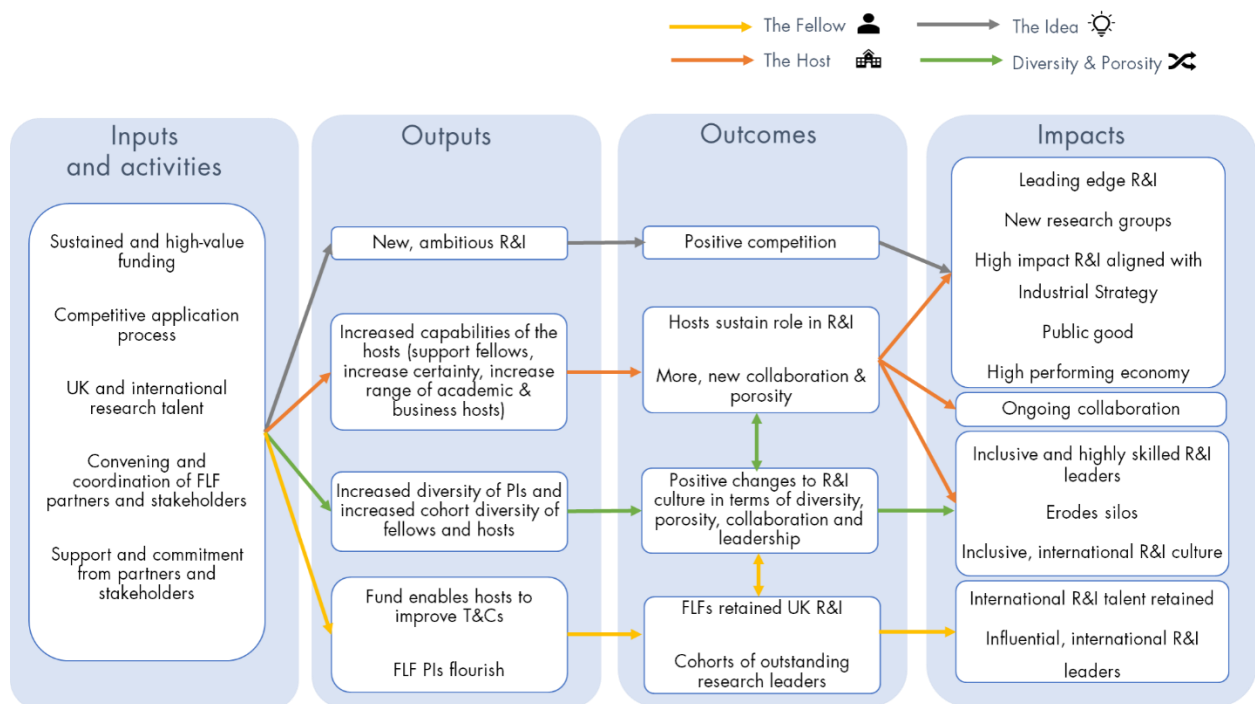


In the sections below, we briefly summarise the four causal pathways of the FLF as captured by the simplified ToC.

## 2.1.1.   The fellow

The fellows are the primary recipients of the FLF intervention and the main unit of analysis in our evaluation. They are one of the agents for change- through the support provided by the FLF programme, they should be enabled to work in multidisciplinary projects, developing and nurturing skills and partnerships, creating culture change at host institutions and forging better links with industry. The anticipated impact is that the FLF facilitates knowledge and its exchange, and cultivates outstanding and diverse cohorts of research leaders, leading to a more inclusive and multidisciplinary R&I ecosystem and better retention of R&I talent in the UK.

### 2.1.2. The host

The hosts are an agent for change that play a crucial role in managing the fund and providing institutional support for the fellows. This includes allowing fellows to spend their working time on the fellowship, financially contributing to the cost of the fellowship (which eventually rises to 80% contribution) and for fellows in academia, awarding a UK-based permanent position at the end of the fellowship. The FLF has the ability to create a diverse cohort of hosts across various sectors and industries and augment their capabilities by creating more certainty around support for the fellows. Through the FLF, the hosts are then able to provide better support for fellows' careers and ambitious projects, build on fellows' achievements to develop practices within and beyond their organisations, and create new networks. The hosts themselves benefit significantly as well: the fellow can help the host develop new expertise, products, and services, thereby creating more economic opportunities. Moreover, in conjunction with the fellows' causal pathways, a multiplier mechanism is enacted whereby a culture of good leadership, networking, and training support enables others to succeed and creates positive feedback. In the wider environment, this leads to more ground-breaking R&I, a more inclusive and diverse R&I culture, and greater retention of R&I talent.

### 2.1.3. The idea

The 'idea' is the amalgamation of the change being created through the FLF programme in the wider ecosystem within with the hosts and fellows are situated. It is anticipated that the idea will lead to new and ambitious R&I projects and promote positive competition and collaboration. The realisation of the idea is anticipated to lead to new research groups, spin-offs, new knowledge and/or new products, and ultimately-economic and social change and impactful and cutting-edge R&I aligned to government strategies and objectives.

### 2.1.4. Diversity and porosity

The design of FLF intends to encourage diversity, collaboration, and porosity. Moreover, FLF is positioned to attract people, hosts and partnerships that also value diversity and are interested in multi- and interdisciplinary research. The diversity and porosity pathway is unique in its ability to affect attitudinal and behavioural change, facilitate networking, exchange of knowledge, and the breaking down of siloes; thereby creating diverse and innovative partnerships, projects, and leaders. This is a cross-cutting pathway that is linked to the previous three pathways of the fellow, the idea and the host.

### 2.1.5. ToC assumptions and context

Although the FLF ToC for this programme as it currently stands does not list any assumptions or context within which this programme is being delivered, we are cognisant that a range of factors will have an impact on the realisation of the intended programme logic. Firstly, it is important to recognise that the ToC is underpinned by a range of implicit assumptions about the expected behaviour of entities across different causal pathways. Such assumptions include hosts' willingness to engage and support fellows, businesses changing their mode of operation, and fellows staying at a given host institute for a reasonable length of time etc.

The implementation of FLF programme to date has also taken place in a specific context including exogenous shocks resulting in significant changes in the economic and societal landscape. Brexit, for example, has impacted the R&I sector through the limited uptake of Horizon 2020 grants, with UK applications falling by 40% in 2018 when compared to previous years.[20] Covid-19 is also a relevant contextual factor which will have impacted host institutions in their financial stability, retention of the research workforce and prioritisation of research projects. Such changes will necessarily have impacted the ability of the FLF portfolio to deliver upon its intended aims. FLF is also being implemented in the context of a changing political and research landscape. Emerging national priorities, such as the levelling up agenda and EDI, have gained increasing precedence over the lifespan of the programme.

**Table 1: Summary of key ToC assumptions and contextual factors**

| Key ToC assumptions | Key ToC contextual factors |
|---|---|
| • Fellows are committed to using support provided by FLF to pursue novel interdisciplinary research, build skills, foster collaborations and partnerships and progress towards leadership positions in their fields, in line with the programme's objectives.<br><br>• Host institutions will support fellows in the above pursuits by enabling sufficient time for research, contributing to the cost of fellowships, and offering longer-term employment opportunities. In turn, hosts will also recognise opportunities for institutional growth tied to the fellowship, including development of new research areas, new collaborations, and new economic opportunities.<br><br>• Support provided to fellows will lead to the advancement of new knowledge and ideas, which in turn will underpin the creation of new knowledge products, research groups, spin-offs and commercialisation opportunities, and broader, longer-term societal change.<br><br>• Prioritising EDI in the allocation of FLF awards will have knock-on effects in terms of the diversity and porosity of fellows' research teams and broader institutional cultures, thereby leading to a more equal, diverse and inclusive R&I workforce. | • The UK's departure from the EU has led to a decline in UK applications for EU Horizon 2020 funding grants, resulting in increased applications to and competitiveness of UK funding schemes.<br><br>• Covid-19 has impacted the ability of fellows to conduct research and on the financial stability of host institutions, including their ability to retain the research workforce and their approach to prioritisation of research projects.<br><br>• The evolving policy landscape has placed increased emphasis on areas such as EDI and levelling up. |

The extent to which underpinning assumptions have held up, and the impact of contextual factors will be important to consider and test during the course of the evaluation. As explained further in Chapters 3 and 5, these assumptions and contexts will be evaluated as part of our theory-based realist approach which is focussed on understanding the context and the mechanisms surrounding the realisation of anticipated outcomes.

---

[20] Royal Society, 'Brexit uncertainty harming UK science', 16 October 2019, https://royalsociety.org/news/2019/10/brexit-uncertainty-harming-UK-science/

## 2.2.    Kirkpatrick model

The Kirkpatrick model is an analytic framework used traditionally for training interventions and has been modified for the evaluation to expand its uses.[21] The modifications we have made are primarily introducing the ROI level in the pyramid, mapping each of the existing levels (i.e. Reaction, learning etc.) to our process and impact evaluations, and ascribing the relevant outputs and outcomes to these levels as listed in the evaluation framework presented in Chapter 5.

**Figure 7: Analytical framing for the evaluation drawing on the Kirkpatrick model**
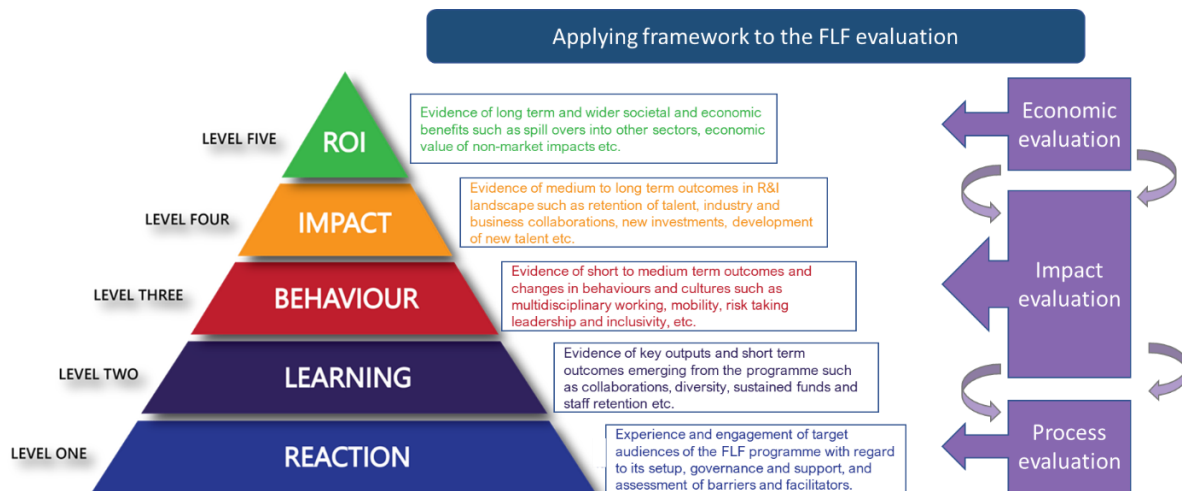
The Kirkpatrick model evaluates the efficacy and impact of interventions across a series of 'levels', which traditionally are: 'reaction' (have recipients found the intervention relevant, useful and engaging), 'learning' (have recipients acquired knowledge, skills and attitudes as intended by the intervention), 'behaviour' (has recipients' longer-term behaviour changed as a result of the intervention) and 'results' (has the intervention achieved its intended outcomes). For the purposes of our evaluation, an adapted version of the Kirkpatrick model will be used to map the impact of the FLF against five hierarchical levels (L1-L5), with each level in turn relating to a different aspect of the processes, impacts and economic outcomes of the FLF programme. For instance, L1- 'Reaction', will provide the framing to assess and synthesise the measures of success for the processes underpinning the FLF set up, governance and recipient perceptions to glean whether L1 has been achieved successfully to influence L2- 'Behaviour'. Similarly, the assessment of outputs and outcomes through various indicators and data collection methods discussed in Chapter 5 will provide a sense of whether the FLF programme as a whole has progressed to L2/L3 etc. at a given point in time. Figure 7 above provides an illustration of how the Kirkpatrick model will be applied to our evaluation of FLF.

The Kirkpatrick framework will thus create a 'hierarchy of impact' to help understand the extent to which the programme as a whole has traversed its trajectory of impact at a given point in time. While the FLF

---

[21] See for example Jones C, Fraser J, Randall S. The evaluation of a home-based paediatric nursing service: concept and design development using the Kirkpatrick model. Journal of Research in Nursing 23(6) (2018):492-501. https://doi.org/10.1177/1744987118786019; Heydari, M.R., Taghva, F., Amini, M. et al. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. BMC Res Notes 12, 388 (2019). https://doi.org/10.1186/s13104-019-4421-y

ToC provides an underpinning framework for assessment against the envisioned inputs to impacts and answering the evaluative questions, the Kirkpatrick model provides added value as an overarching analytical lens drawing together the evidence across the evaluation.

# 3. Evaluation approach

This chapter sets out the overarching evaluation approach through which we will collect evidence to assess the FLF against the analytical frameworks described in Chapter 2. Following the recommendation of the IES scoping and feasibility report, and the specifications of the ITT, our evaluation approach will comprise three types of evaluation: a process evaluation; an impact evaluation comprising a combination of QED and realist TBE; and an economic evaluation. The sections below provide more information on the aims of each type of evaluation.

**Figure 8: Overarching evaluation approach**



## 3.1. Process evaluation

The process evaluation will seek to understand the extent to which FLF's organisational targets have been achieved, including whether programme management, support and structures have been fit for purpose; and to develop an understanding of barriers and facilitators faced by the programme in relation to its setup and implementation.[22] The process evaluation will use a theory-based mixed-methods approach guided primarily by the evaluation questions derived from the ToC, with Kirkpatrick model L1 and L2 used as an

---

[22] In this phase of the evaluation, we will focus on understanding the set-up, operational delivery, and implementation of the FLF programme, rather than the processes through which impacts have been achieved. The processes underpinning impacts will be explored further as part of the impact evaluation where we will adopt a context mechanisms outcomes (CMO) analytical framing allowing for a better understanding of linkages between process and impact. Moreover, we will utilise a cause consequence analysis to assess if processes are likely to impede or facilitate impact.

additional lens for data collection and analysis. The process evaluation methods will comprise document review, management information analysis (in addition to that already undertaken for this framework report), key informant interviews, process mapping, cause consequence analysis, case studies and comparator case studies.

## 3.2. Impact evaluation

The impact evaluation will provide an assessment of the outputs, outcomes and impacts outlined in the ToC. The evaluation will focus on understanding the extent to which FLF has achieved shorter-term outputs and outcomes, while also considering whether the programme is on track to achieve longer-term desired outcomes and impacts. The impact evaluation methods will comprise a review of internal UKRI data and reporting, a review of secondary data, career tracker surveys, a host survey and key informant interviews.

The impact evaluation will comprise two related but distinct strands: (i) a counterfactual QED where we seek to understand the **causal impact of the FLF quantitatively, focussing on short-term outputs,** and (ii) a theory-based approach to capture more qualitative outputs and outcomes, why and how they have occurred, and whether the FLF is on track to achieve a sustained longer-term impact. Given the nature of the FLF programme and the outcomes anticipated, for example supporting multidisciplinary research and developing leadership, it is evident that a qualitative approach will be more beneficial in drawing out the nuances of what this means and how these outcomes vary in their manifestation across diverse sectors and disciplines. We therefore propose that the QED should be considered a limited component of a broader contribution story of the FLF which will be strengthened through qualitative research. More information on these two strands is provided in the sections below.

### 3.2.1. Quasi-experimental design (QED)

The purpose of the QED evaluation is to estimate the causal impact of the FLF programme on observable outcomes. It will do so by generating an *ex post* counterfactual group of unsuccessful awardees, who are otherwise similar to the successful applicants (the treatment group) in all observable ways and comparing differences in outcomes between the two groups. The key research question that will be addressed through the QED is whether FLF fellows achieve the outputs anticipated to a higher degree compared to the counterparts. We are not intending to compare the FLF with other fellowships given the challenges of generating a counterfactual (see Chapter 4). To assess the added value of the FLF over other interventions, we will utilise a qualitative approach, as is further described in Section 3.2.2 and Chapter 5 below.

The 'gold-standard' in identifying causal impact is generally through a Randomised Control Trial (RCT), where treatment (the awarding of the FLF in this case) is randomly assigned to eligible participants, creating a control, or counterfactual, that is otherwise equivalent to the treatment group. RCTs can assign causal impact as they allow us to control for any changes in outcomes that would have occurred without treatment anyway, and isolate the impact of the fellowship itself. QED evaluation, however, tries to non-experimentally identify a counterfactual that is as similar to the treatment group in all observable characteristics. By identifying an appropriate counterfactual who it is estimated would follow the same outcome trajectory as the treated individuals, had those individuals not been treated, the QED can then

attribute any differences in outcomes to the treatment (the FLF intervention) itself. This enables causal estimation of the receipt of the fellowship on individual-level outcomes.

The QED can only estimate the causal impact of fellowship receipt on quantifiable, individual-level outcomes. Even though outcome metrics may exist for other units of analysis as outlined in the ToC, such as the host institution or collaborators, we could not credibly and causally attribute any differences in outcomes at these broader levels to the receipt of fellowship. It is unlikely that the necessary assumptions to attach a causal interpretation hold at levels other than that at which the counterfactual is generated. As such, given treatment and counterfactual generation both occur at the individual level, we only attach causal interpretation to differences that occur in researcher-specific outcomes.

In Chapter 4, we set out further considerations relevant to the QED evaluation and the composition of the counterfactual group. In Chapter 5, we present our impact evaluation framework including the specific observable outcomes for which we will seek to collect data for the QED evaluation.

### 3.2.2. Theory-based evaluation (TBE)

A realist TBE will form the second strand of our impact evaluation. We propose a realist TBE over other TBEs due to the emphasis placed on the context and mechanisms surrounding outcomes, which will be crucial in eliciting learnings for the programme. While a QED evaluation will provide quantitative measures on a limited set of FLF's outputs, it will not factor in qualitative outputs and outcomes, nor proxies for longer-term impact. The TBE will enable consideration of these wider forms of impact, including the diversity of impacts across sectors and disciplines. This is particularly important given the business and academic settings of the FLF and the range of STEM and AHSS disciplines involved. Our TBE will also seek to gather qualitative evidence regarding the extent to which FLF has provided added value over other similar interventions. The incorporation of realist approaches into our TBE will create scope for understanding why certain outcomes have occurred, exploring the specific contexts in which they have occurred, and surfacing evidence regarding the unintended consequences of the FLF. We will utilise the Context Mechanism Outcomes (CMO) analysis framing to undertake the realist evaluation, as further described in Chapter 5.

## 3.3. Economic evaluation

The final phase of the evaluation will comprise an economic evaluation of the FLF, contingent upon undertaking an economic feasibility assessment to assess the viability of a cost benefit analysis of FLF and to explore potential approaches to assessing value and effectiveness.

Broadly, there are two potential options for the economic appraisal of FLF: a direct cost benefit analysis of the FLF investment; or an exploration of the opportunity cost of the FLF (for example, by examining the potential benefits of running the programme differently or funding an alternative prospect through a discrete choice experiment). The development of a direct cost benefit analysis for FLF will require a review of the outputs and outcomes measures as agreed and quantified in the impact evaluation to assess whether they are monetizable. It will also require a review of all the qualitative data collected to assess whether there is anything else that could be quantified and monetised. Where possible, we will explore use of multipliers and look to expand the economic impact measures to consider labour market effects and spillovers (while

also recognising limitations relating to the timing of the evaluation at a relatively early stage of implementation of the FLF). Beyond a cost benefit analysis, some elements of cost effectiveness or value may be possible to discern through a focussed assessment of additionality. This would be a qualitative approach assessing whether the investment was effective in achieving its outcomes or whether the outcomes would have been achieved regardless of the intervention. These options will be explored through the feasibility assessment.

# 4. QED considerations and scoping of the counterfactual

This chapter sets out considerations for the QED evaluation based on the evaluation team's further assessment of the IES scoping and feasibility study[23] and the FLF management information shared by UKRI. The IES scoping and feasibility study focused principally on two forms of QED: Propensity Score Matching (PSM) and Regression Discontinuity Design (RDD), and ultimately recommended the PSM approach. Following further consideration, we agree with their summary and recommend the adoption of PSM. These approaches largely differ in respect to counterfactual generation. PSM generates a counterfactual by modelling the likelihood of being treated- based primarily on the Sift score, and generating a group of unsuccessful applicants that have a similar likelihood of being treated as the treatment group. RDD instead relies on identifying the threshold for treatment and comparing those either side of the threshold. While we can identify a threshold using Interview Scores, there is unlikely to be a sufficient sample size around the threshold to detect an effect size.[24] PSM allows us to generate a counterfactual with a similar likelihood of being treated, at the application stage rather than the interview stage, thus allowing us to retain a wider pool of individuals in both counterfactual and treatment.[25]

We propose using RDD, which uses a narrower counterfactual group, as a robustness check on the PSM results. RDD identifies a threshold and uses this threshold to generate a pseudo-random counterfactual by focussing on those around the threshold. However, by reducing the sample to just those on either side of the cut-off, the RDD analysis will likely be underpowered and struggle to detect a significant effect. As it is likely to be underpowered, RDD results are best used as a robustness check, rather than as the primary analytical approach. If the estimated effect of receiving a fellowship under RDD is broadly similar in estimated magnitude and direction to that estimated using PSM, this would suggest that the findings are robust to methodological approach and counterfactual generation.

We suggest that the counterfactual generation is revisited once we have a set of outcome variables, to establish whether the outcome variables differ substantially with respect to fellowship timing (particularly whether pre- or post-pandemic) or with respect to researcher discipline. If either of these hold, then the counterfactual sample may need to be restricted further, or alternatively, additional controls used in a pooled analysis. Any restrictions on the counterfactual made now or necessarily put in place after a review of the outcomes are specific to the QED analysis. The TBE analysis need not adopt such a narrow approach to

---

[23] Emma Pollard et al., "UKRI Future Leaders Fellowships Evaluation Scoping & Feasibility Study."

[24] Emma Pollard et al.

[25] RDD could also be broadened to include the wider applicant pool, but this would require the adoption of a fuzzy threshold rather than a sharp threshold. This was considered and ruled out by Pollard et al and we agree with their assessment.

the sample of interest. Instead, the TBE can adopt purposive sampling to analyse those fellows and outcomes that the QED cannot adequately address.

## 4.1. Propensity Score Matching (PSM)

PSM compares successful applicants with unsuccessful applicants who have a similar likelihood of being awarded a fellowship, based on the set of observable characteristics available at application. It generates the likelihood of being awarded a fellowship by modelling the probability of success, or the propensity of being treated, as a function of relevant observable characteristics. This is usually done using a logit or probit regression model.

PSM tries to mimic *ex post* an experimental control group through the creation of a counterfactual group from among non-treated individuals, such that the counterfactual is similar to the treatment group in all relevant observable characteristics.[26] It ensures the distribution of covariates[27] in this comparison counterfactual is the same as the distribution in the treatment group.[28] A causal interpretation can be attributed to any differences in the outcomes between the treated and counterfactual if the counterfactual generation meets the following two assumptions: i) treatment depends only on the included observable covariates, otherwise known as the *selection on observables* assumption, and ii) that treated individuals can be matched with a counterpart who has a similar likelihood of being treated, or the *common support* assumption.[29] The validity of PSM as an approach to causal estimation depends on the validity of these assumptions.

With regards to the FLF, this first assumption equates to the following: the awarding of a fellowship is based only on the observable characteristics of an applicant. Generally, this assumption would not hold if there are any unobserved characteristics, such as drive or ability, that determine whether or not an individual is successful in being awarded a fellowship. Given that we have access to both Sift scores and Interview scores[30]- which are ordinal scores that reflect the beliefs of evaluators about the overall suitability of a candidate and their research proposal for the FLF, including these other characteristics that are less readily observed, this assumption is likely to hold. This is particularly true of Interview scores, where it is expected that these more traditionally unobservable characteristics can be observed when evaluators meet the candidates in person. If it is not observable at that stage, then individuals cannot be selected on it, and the *selection on observables* assumption would still hold.

---

[26] Richard Blundell, Lorraine Dearden, and Barbara Sianesi, "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey," *Journal of the Royal Statistical Society: Series A* 168, no. 3 (2005): 473–512.

[27] A covariate can be defined as any independent variable (in this instance individual characteristics and application characteristics) that can influence the outcome (here, the likelihood of being awarded an FLF) but is not of direct interest. Examples of covariates could include gender or age, but also Sift score.

[28] Zhong Zhao, "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics and Monte Carlo Evidence," *Review of Economics and Statistics* 86, no. 1 (2004): 91–107.

[29] Blundell, Dearden, and Sianesi, "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey."

[30] The Sift score is the score assigned to an individual's application by the evaluating panel that reflects their evaluation of the proposal and the researcher against the assessment criteria. The Interview score is the updated score assigned to an individual's application at interview, should they be invited to interview, based on their revised evaluation of the proposal and researcher against the assessment criteria.

If there are some treated individuals for whom no counterfactual individual can be found (normally those with the highest probability of being treated), then these individuals may need to be removed from the treatment group before estimation by PSM can be undertaken. Similarly, untreated individuals who have the lowest probability of being treated are also unlikely to be included in the counterfactual. This is a more problematic assumption in that it is unlikely that those with the highest Sift and Interview scores are not successful and unlikely that those with the lowest score are ever successful. However, it can also be readily fixed by restricting the analysis to just those propensity scores for whom there are both successful and unsuccessful candidates. However, this may also necessitate a reinterpretation of any measured effect to the average treatment effect of the treated within the *common support*.

The difference between PSM and other regression techniques can be subtle. It has been demonstrated that ordinary least squares (OLS) regression is biased even when treatment is purely a function of observable characteristics, if either there is non-overlapping support of the observables or there is mis-weighting on the common support.[31] Under either of these assumptions, the empirical distribution of observables across the treated and non-treated individuals are not the same and can result in bias estimates of the average treatment effect on the treated. However, it is important to note that PSM cannot overcome omitted variable bias- the most easily recognised of the forms of bias in OLS regression techniques, as the presence of omitted variable bias in OLS also ensures that the first PSM assumption- *selection on observables*, cannot be satisfied.

In determining the suitability of PSM, we have focussed on using the pool of unsuccessful applicants as the counterfactual. An alternative to using non-successful applicants as a counterfactual is to use individuals in receipt of other fellowships as a counterfactual. Whilst these individuals may be more likely to follow a similar trajectory to FLF fellows over the course of their career, and so may seem like a more appropriate control group, selecting them does not enable us to quantitatively estimate what the effect of receiving the FLF is on researchers. Instead, it would give an estimate of the additional effect of receiving the FLF instead of some other early career fellowship. Furthermore, it is more difficult to justify one of the key assumptions of PSM, that of selection only on observables, using other fellowship recipients as the counterfactual. The advantage of using the unsuccessful applicants is that by virtue of the scoring system used at the application and interview stage, all individuals are given a numerical score for suitability based only on observables and it is this numerical score on observables that ultimately determines whether they progress or are selected, ensuring the propensity score model satisfies the *selection on observables* assumption.

## 4.2. Generating the counterfactual

In order to meet the requirements of both PSM assumptions, we need to first ensure that we are using all relevant observables in generating the propensity score. Given that the award of a fellowship is based on the subjective opinions of evaluators on the suitability of the candidate and their research proposal, we need to ensure that there is some metric that summarises these subjective evaluations. Fortunately, both the Sift score and Interview score are suitable candidates for such a metric as they provide a single ordinal metric

[31] James J Heckman, Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies* 65, no. 2 (1998): 261–94.

that sums up the beliefs of the panel of evaluators on the overall suitability of a research proposal and candidate.

Table 2: Number of applicants at sift, interview and award stage by round

|  | Rounds 1 & 2 | Rounds 3 & 4 | Rounds 5 & 6 |
|---|---|---|---|
| Sift Scores | 758 | 742 | 1398 |
| Interview Scores | 256 | 326 | 344 |
| Awarded | 127 | 192 | 180 |
| STEM Awardees | 104 | 159 | 142 |
| Social Studies and Humanities Awardees | 23 | 33 | 38 |

Source: RAND Europe analysis

However, given only a limited number of candidates are invited to interview, most candidates are only assigned a Sift score. Table 2 shows that approximately 34%, 44% and 25% of applicants are invited to interview in Rounds 1 and 2, Rounds 3 and 4 and Rounds 5 and 6 respectively. And of those invited to interview, approximately 50% or more are awarded a fellowship. Given the further restrictions on sample size necessary to meet the *common support* assumption, it is unlikely that we would have a large enough counterfactual to enable analysis using Interview scores. Although this is necessary, it does have some implications for the assumption of *selection on observables* if there are some characteristics that are observable at interview but not at point of application. We proceed for now on the basis that this is not the case, such that including Sift scores is sufficient to ensure this assumption holds; but can restrict the analysis to candidates invited to interview as robustness check using RDD or otherwise to see how this alters estimates.

Table 2 also demonstrates another restriction, about sample size- which is that the number of STEM awardees is around four times the number of Social Studies and Humanities awardees, if not larger. The number of business awardees is even lower than AHSS. This is not as much a problem for the counterfactual creation itself, but rather a problem with regard to possible outcome metrics. A review of possible outcome metrics suggests that there are unlikely to be many that are applicable to both academic and business awardees. Therefore, in our counterfactual creation, we restrict ourselves to academic candidates only. However, even outcome metrics that exists for academic fellows as a whole may follow different trends depending on discipline. For instance, the number of collaborators and types of publications are often different across disciplines, as can be citation metrics. As we do not yet have access to the possible outcomes, we cannot empirically evaluate this claim so have for the moment generated the counterfactual on all academic awardees. However, any substantive differences in trends in outcome metrics across different disciplines would require a further sample restriction for counterfactual generation. To ensure that we still get a complete picture of the effect of FLF on all recipients, the TBE can be targeted to those who are more difficult to analyse quantitatively to ensure this gap is filled. Any reduction on sample necessary for QED does not mean the TBE sample must be similarly restricted.

Further restrictions in sample size are necessarily required to ensure the *common support* assumption – is satisfied. As discussed in detail in Annex C, this assumption is not satisfied using the full sample of successful

and unsuccessful applicants. There are several successful applicants whose likelihood of being treated as calculated using the final PSM model, is so high (approximately 1) that there are no unsuccessful individuals who have the same likelihood of being treated. There are also a large number of unsuccessful applicants whose likelihood of being treated is so low that there are no successful individuals who have the same likelihood of being treated. PSM necessarily excludes both this upper and bottom tail from counterfactual generation and analysis to ensure this assumption is satisfied.

The ability to assign causality in QED methods crucially requires that we can argue that the only difference between treatment and counterfactual is the assignment to treatment. With regards to outcomes, however, these can be substantially changed by macroeconomic shocks, two of which occurred during the period over which the FLF rounds were conducted and could substantially alter the effect of the fellowship on researcher outcomes. These two shocks are: i) the UK formally leaving the European Union ('Brexit') and ii) the Covid-19 pandemic. Both of these happened at the beginning of 2020, between rounds 4 and 5. Particularly with regards to the effect of the pandemic on the treatment effect, it seems likely that the effect on subsequent career trajectories depends on when during the course of the fellowship or pre-fellowship period did the first and second lockdowns occur. We have two possible options to control for these shocks: the first is to pool applicants, using the bias-correction techniques for split cohorts[32] and adding in controls for the macroeconomic shocks; and the second is to do so by sub-dividing the treatment and counterfactuals into different subsets based upon year of application and analysing each subset separately to avoid introducing bias into the treatment effect.

We propose using both these techniques in the impact evaluation but given the second option places further sample size restrictions on the counterfactual, we present the exploratory counterfactual generation by sub-setting on year of application. We break the sample into three sub-samples: those who applied during 2018 (Rounds 1 and 2), those who applied during 2019 (Rounds 3 and 4) and those who applied after the onset of the pandemic (Rounds 5 and 6).

Table 3 shows the differences in proportions or mean between successful and unsuccessful applicants across a number of different covariates, for each of the three subsamples. As expected, there are significant differences in the Sift scores between successful and unsuccessful applicants across all three sub-samples. We also notice that there are some significant differences between successful and unsuccessful applicants across the other observable covariates as well, although these differences are not significant in all rounds. Given we wish to use the same model for calculating propensity scores across all sub-samples, we control for all covariates that are significant in any individual round in the estimation of propensity score matching. Additionally, given the decision to leave the European Union caused a change in ease of access to the UK labour market for non-UK academics during this period, we also control for UK citizenship. We evaluate

---

[32] Brantly Callaway and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (December 1, 2021): 200–230, https://doi.org/10.1016/j.jeconom.2020.12.001.

the success of the model in generating a counterfactual by comparing the similarity of the counterfactual to the treatment group across covariates.[33]

Table 3: Proportion or mean of successful and unsuccessful applicants across covariates by round

| | Female | Under 40 | White | UK Citizen | Sift Score |
|---|---|---|---|---|---|
| Rounds 1 and 2 | | | | | |
| Unsuccessful | 0.37 | 0.80 | 0.77 | 0.48 | 4.41 |
| Successful | 0.43 | 0.85 | 0.76 | 0.54 | 8.31*** |
| Rounds 3 and 4 | | | | | |
| Unsuccessful | 0.35 | 0.80 | 0.69 | 0.26 | 4.55 |
| Successful | 0.47*** | 0.83 | 0.84*** | 0.21 | 8.32*** |
| Rounds 5 and 6 | | | | | |
| Successful | 0.40 | 0.78 | 0.73 | 0.26 | 5.19 |
| Unsuccessful | 0.43 | 0.85** | 0.79 | 0.23 | 8.5*** |
| *, **, *** denote p-values of 0.1, 0.05 and 0.01 respectively, when successful applicants are compared with unsuccessful applicants | | | | | |

Source: RAND Europe analysis

Using PSM, we have been able to generate a counterfactual that looks sufficiently similar to the treatment. The details of the PSM model used and the various matching methods considered are detailed in Annex C. The best of these alternatives, with regards to maintaining sample size and generating a well-balanced counterfactual, is coarsened exact matching. Coarsened matching effectively coarsens covariates by grouping covariates into bins or subgroups and matching on these. Coarsened matching, like nearest neighbour matching, considers the full counterfactual. However, unlike nearest neighbour matching, it does not force 1:1 matching and instead allows some individuals to be matched multiple times if that improves the balance of covariates across counterfactual and treatment. The results for this approach are given in Table 4 and the metrics to assess the match provided in Annex C.

---

[33] Checking for significant differences across the covariates is not necessarily informative in evaluating the success of counterfactual generation. Standard metrics have been used to evaluate these approaches and further details can be found either in the annex or upon request.

**Table 4: Comparing treatment and PSM-generated counterfactual – coarsened matching**

|  | Sample Size[a] | Female | Under 40 | White | UK Citizen | Sift Score |
|---|---|---|---|---|---|---|
| Rounds 1 and 2 | | | | | | |
| Treatment | 89 | 0.39 | 0.90 | 0.81 | 0.54 | 8.63 |
| Counterfactual | 91.01 | 0.39 | 0.90 | 0.81 | 0.54 | 8.62 |
| Rounds 3 and 4 | | | | | | |
| Treatment | 153 | 0.39 | 0.88 | 0.86 | 0.53 | 8.64 |
| Counterfactual | 61.06 | 0.39 | 0.88 | 0.86 | 0.53 | 8.58 |
| Rounds 5 and 6 | | | | | | |
| Treatment | 158 | 0.44 | 0.89 | 0.82 | 0.55 | 8.98 |
| Counterfactual | 129.96 | 0.44 | 0.89 | 0.82 | 0.55 | 8.92 |

[a] For counterfactual, this is the effective sample size which adjusts the raw sample size to take into account that some individuals may be matched more than once.

Source: RAND Europe analysis

Coarsened matching is not the only approach, with some others considered in Annex C, and we remain open to using another matching approach in the analysis. The above results are indicative and suggest that it is possible to use PSM to establish the causal effect of the FLF on researcher outcomes. However, we would suggest we re-visit the counterfactual generation once we have a set of outcome variables; so as to establish whether the outcome variables differ substantially with respect to fellowship timing (particularly whether pre- or post-pandemic) or with respect to researcher discipline - so that we can finalise the subsets over which we wish to match.

We will also consider comparing results from PSM with results generated from RDD, to evaluate the robustness of the results to differences in QED approach. The advantages of PSM over RDD are principally with regard to sample size – it allows for a larger counterfactual and thus improves the ability of our model to detect differences in effect size. Given the relatively small number of individuals invited to interview, this would result in a substantial fall in sample size. Even with the restrictions in sample size necessitated by the *common support* assumption, PSM would still have a higher sample size that that under RDD with a sharp threshold. However, it is likely that the relative gains of PSM over RDD are lower than first thought under the feasibility assessment conducted by the IES.

# 5. Evaluation framework

This chapter sets out the framework through which the evaluation of the FLF will be operationalised. The chapter comprises two distinct but interrelated frameworks: a process evaluation framework and an impact evaluation framework. Presented in tabular form, each framework delineates evaluation questions (EQs), derived from the FLF ToC, which our evaluation seeks to answer. For each EQ, the frameworks delineate key metrics and proposed data sources that will be used to collect evidence to enable us to answer the EQ. Where appropriate, both frameworks also highlight the alignment of metrics to levels of the Kirkpatrick model. Alongside process and impact evaluation frameworks, the chapter also provides further detail on the data collection methods and analytical approaches that will be used for each type of evaluation.

## 5.1. Process evaluation framework

As highlighted in Chapter 3, the process evaluation will seek to understand the extent to which FLF's organisational targets have been achieved; whether programme management, support and structures have been fit for purpose; and to develop an understanding of barriers and facilitators faced by the programme in relation to its setup and implementation. This will be followed by developing an understanding of any causal mechanism through which processes could affect impacts of the programme. The approach follows the Medical Research Council (MRC) process evaluation framework[34] in emphasising causal pathways and we propose to review the link between process and impact through a cause consequence analysis (see Section 5.1.1). Table 5 below sets out our process evaluation framework. The EQs included within the framework have been selected for their alignment to process themes identified by the evaluation team derived from inputs, activities and outputs within the FLF ToC. The sections beneath Table 5 describe in more detail the data collection tools and analytical approaches that will be used to undertake the process evaluation.

---

[34] Graham F Moore et al., "Process Evaluation of Complex Interventions: Medical Research Council Guidance," *BMJ : British Medical Journal* 350 (March 19, 2015): h1258, https://doi.org/10.1136/bmj.h1258.

## Table 5: Process evaluation framework

| ToC process themes (derived from inputs, activities and outputs) | Evaluation question | Metric | Proposed data sources | Kirkpatrick level (as applicable) |
|---|---|---|---|---|
| Performance | To what extent have the FLF applicant, awardee and financial, targets been met? | Number of applicants against target | Management information analysis | |
| | | Number of fellows awarded against target broken down by sector, discipline, geography and nationality | Management information analysis | |
| | | Amount of funding provided to fellows/hosts against target broken down by sector, discipline and geography | Management information analysis | |
| Diversity | To what extent and how have FLF processes supported multidisciplinarity in the fellowship scheme? | Inclusiveness and flexibility of the application criteria | Document review; key informant interviews | |
| | | Transparency and design of the programme and award criteria | Document review; key informant interviews | |
| | | Diversity make-up of funding panel/award panel | Document review; key informant interviews | |
| | | Comprehensiveness of channels used for communicating scheme/pre-launch engagement etc. | Document review; key informant interview | |
| | | Number of fellows awarded against target broken down by sector, discipline, geography and nationality | Management information analysis | |
| | | Amount of funding provided to fellows/hosts against target broken down by sector, discipline and geography | Management information analysis | |
| Value | To what extent is the premise of the FLF scheme unique in the sector? | Nature of interactions and support given to hosts (e.g. certainty created by funds) | Key informant interviews; document review | 1-Reaction |
| | | Nature of interactions and support provided by hosts to fellows | Key informant interviews; document review | 1-Reaction |
| | | Assessment of networks and platforms available to fellows | Key informant interviews; document review | 1-Reaction |
| | | Changes (if any) in host values, T&Cs in response to scheme | Key informant interviews; document review | 2-Learning |

| | | Assessment of cross-fellow and host interactions (e.g. between business and academia) | Key informant interviews; document review | 2-Learning |
|---|---|---|---|---|
| Implementation | To what extent has FLF delivered effective post award management to support the professional development of the fellows? | Nature of interactions and support given to hosts (e.g. certainty created by funds) | Key informant interviews; document review | 1-Reaction |
| | | Nature of interactions and support provided by hosts to fellows | Key informant interviews; document review | 1-Reaction |
| | | Assessment of networks and platforms available to fellows | Key informant interviews; document review | 1-Reaction |
| | | Review and assessment of post-award requirements made on the fellows | Key informant interviews; document review | 1-Reaction |
| | What, in practice, is felt to be working more/less well regarding the delivery of the FLF? | Perception of effectiveness and value by hosts and fellows | Key informant interviews | 1-Reaction |
| | | Case studies exemplifying delivery facets that have worked well and where improvements are required* | Key informant interviews; document review | 2-Learning |
| | What are the unexpected barriers or facilitators to FLF processes and the delivery of the anticipated outcomes, if any? | Perception of barriers and enablers by hosts, UKRI and fellows based on delivery and post-award management phase | Key informant interviews | 1-Reaction |
| | What lessons are there for future rounds/similar schemes? | Assessment of process improvements and learning between each cohort and at delivery phase | Key informant interviews | 2-Learning |
| | | Assessment of factors that hindered/facilitated programme targets | Key informant interviews; document review | 1-Reaction |

* Our approach to process case studies is described further in Section 5.1.1 below.

### 5.1.1.  Process evaluation analytical approaches

The process evaluation will involve four analytical approaches: process mapping, a cause consequence analysis, process case studies and comparator case studies. These various analytical approaches will help to build a 'living' picture of FLF processes, including examples of programme processes in action, thereby helping to comprehensively answer the EQs. The four analytical approaches are described in more detail below.

## Process mapping

We will synthesise the data collected through document review, management information analysis and key informant interviews to create an 'end to end' process map of the FLF programme. The process map will divide FLF programme processes into a set of stages – from programme design, through funding calls, application, review and decision-making, to post-award and monitoring and evaluation. The exercise will also identify areas of strength and opportunities for improvement in programme processes from the data analysed.

## Cause consequence analysis

As part of the process framework evaluation approach focussing on context and causality, we will conduct a cause consequence analysis (CCA) to assess whether processes that have not worked well could end up affecting the impacts to be realised. CCA is a method for illustrating the possible outcomes arising from a given set of input states or events.[35] For each sub-optimal process or challenge encountered, we will develop a hypothesised consequence on the outcome realisation and underpin this consequence with a proposed assumption of the causal mechanism based on the evidence review conducted prior to the exercise. The CCA, which will produce a consequence tree and causal assumptions map, will provide contextual framing for the assessment of programme impacts later in the evaluation.

## Case studies

We anticipate having enough information to develop three exemplar case studies. As highlighted in Table 5, these case studies are intended to be used to answer the following EQ: *What, in practice, is felt to be working more/less well regarding the delivery of the FLF?* Case studies will be selected purposively and will be used to tease out instances in which processes have worked exceptionally well, and/or where they have not, and where subsequently there may be opportunities for improvement. Here, case studies of the FLF programme may also be linked to comparator case studies conducted from other fellowship programmes. Depending on the nature of the evidence, the case studies could potentially be structured per round of award to explore key themes relating to how the FLF programme has evolved.

## Comparator case studies

Alongside the three case studies of FLF processes, we will also develop two high-level comparator cases studies. The aim of the latter will be to exemplify best practice from other relevant fellowship schemes,

---

[35] National Institute for Health and Care Excellence (NICE), "Evidence Synthesis and Cost-Consequence Analysis - Medical Technologies Evaluation Programme Methods Guide," 2017, https://www.nice.org.uk/process/pmg33/chapter/evidence-synthesis-and-cost-consequence-analysis.

linking these to areas where FLF processes have faced challenges or where potential improvements have been identified. The comparator cases will be identified through information surfaced through key informant interviews, supplemented by a targeted review of the literature focussing on any systematic analyses. Potential examples of comparator programmes include the Henry Dale Fellowships, European Research Council Starting Grants and US National Institutes of Health Early Research Career Development programmes.

## 5.1.2. Process evaluation data collection methods

As shown in Table 5, the process evaluation will rely on three data sources: document review, management information analysis and key informant interviews. These data collection methods are described in more detail below.

### Document review

In addition to the review of the GIAA audit and BEIS documentation already conducted to inform this evaluation framework (see Section 1.5 and Annex A), we will conduct a further review of programme documentation to further develop our understanding of FLF programme processes. Documents to be reviewed will include programme call documentation, review panel documentation and proposal review criteria, and documents relating to FLF marketing, communications and events (e.g. the FLF Conference).

### Management information analysis

While data on FLF applicants and grants has already been reviewed during the evaluation framework stage with a view to scoping the QED counterfactual, we will conduct a further analysis of this management information with a view to answering process evaluation questions regarding the extent to which FLF's organisational targets for applicants, awardees and funding provision have been achieved.

### Key informant interviews

Alongside document review and management information analysis, data on how the programme set up and delivery has been received by hosts, fellows and key stakeholders, and on barriers and enablers to programme implementation will be collected primarily through key informant interviews. To inform the process evaluation we will conduct interviews with the following groups:

- Up to four central operational/delivery staff at UKRI

- Up to 24 FLF fellows

- Up to 10 members of staff at host institutions of FLF fellows

- Up to five unsuccessful applicants to FLF.

FLF fellow and host institution interviewees will be selected using a stratified, purposive sampling approach. For each FLF cohort, we will stratify fellows by criteria such as Research Council, sector (business/academia), subject area, gender and grant size (using ranges). Four fellows from each funding round will then be selected purposively to ensure a mix of interviewees against the stratification criteria (both within each cohort and across the sample of fellows as a whole). From the 24 sampled fellows, 10 host institutions will be selected for interviews with institutional personnel. The 10 institutions will also be

selected to ensure a mix of institution types, both in terms of sector (business/academia) and size. To complement our interviews with FLF beneficiaries (fellows and hosts), we will also conduct up to five interviews with unsuccessful applicants, the aim of which will be to understand their perceptions regarding the process of applying to FLF. The target interviewees will ideally be applicants included within the QED evaluation's counterfactual group. Interview topic guides and analysis coding will be guided by the evaluation questions and L1-2 of the Kirkpatrick model. We will also complement interviews through engagement with FLF fellows at UKRI annual meetings such as the FLF Conference.

## 5.2.  Impact evaluation framework

As described in Chapter 3, the impact evaluation will provide an assessment of the outputs, outcomes and impacts outlined in the ToC, employing a combination of QED and TBE. Table 6 below sets out our impact evaluation framework. Within the impact evaluation framework, overarching EQs are broken down into more specific sub EQs. Where the process framework aligned EQs to process themes based on inputs, activities and outputs, the impact framework aligns EQs to the ToCs four causal pathways of the impact (the fellow, the host, the idea, diversity and porosity). The impact evaluation framework also highlights whether metrics are planned for use within the QED or TBE elements of the impact evaluation.[36] (A more detailed version of the impact evaluation framework delineating the units of data collection (i.e. fellow/host; academia/business; STEM/AHSS) for each metric has also been presented in Annex B.) The sections beneath Table 6 describe in more detail the data collection methods and analytical approaches that will be used in the impact evaluation.

### ToC causal pathways

The fellow

The host

The idea

Diversity and porosity

---

[36] While the categorisation of a metric as either QED or TBE is based on the evaluation team's assessment of data availability at the time of submitting this evaluation framework report, there is a possibility that the collection of data for some QED metrics may encounter challenges during implementation, in which case those metrics may be dropped, revised or incorporated into the broader TBE approach.

## Table 6: Impact evaluation framework

| TOC causal pathways | Overarching EQ | Sub EQ | Metric | QED /TBE | Proposed data sources | Kirkpatrick level (as applicable) |
|---|---|---|---|---|---|---|
| | How has FLF shaped the R&I landscape? | To what extent and how has FLF increased high quality and impactful R&I? | Average field-weighted citation impact of researcher | QED | Dimensions; Scopus | 2-Learning |
| | | | Altmetric attention score | QED | Altmetric | 2-Learning |
| | | | Number/nature of and ability to develop grey literature outputs since start of fellowship | TBE | Researchfish; career tracker surveys | 2-Learning |
| | | To what extent and how does the overall FLF fund support wider government objectives? | Contribution to policy making (including participation in national consultations) | TBE | Researchfish; career tracker surveys | 4-Impact |
| | | | Perspectives on the extent to which FLF has supported wider government objectives, and if so, how | TBE | Key informant interviews | 4-Impact |
| | | | Examples of FLF supporting wider government objectives | TBE | Key informant interviews | 4-Impact |
| | | To what extent and how has FLF increased MIDRI and cross-sector working, for the fellow, fellow's team and/or collaborators? | New and existing collaborations (public/third/academic/corporate) | TBE | Career tracker surveys; Researchfish | 3-Behaviour |
| | | | Number of grants from more than one research council since the start of fellowship | QED /TBE* | UKRI management information; career tracker surveys | 2-Learning |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | Extent of collaborative publications (assessing meta data on multiple disciplines and at least 1 author from a non-HEI) | QED | Scopus; Dimensions | 3-Behaviour |
| | | | Contribution to policy making (including participation in national consultations) | QED/ TBE* | Researchfish; career tracker surveys | 4-Impact |
| | | | Perspectives on the extent to which FLF has increased MIDRI and cross-sector working, and if so, **how** | TBE | Key informant interviews | 3-Behaviour |
| | | | Examples of FLF supporting MIDRI and cross-sector working | TBE | Key informant interviews | 3-Behaviour |
| | | To what extent has FLF increased engagement between industry and academia on R&I activities? | New and existing collaborations (public/third/academic/corporate) | TBE | Career tracker surveys; Researchfish | 3-Behaviour |
| | | | Extent of collaborative publications (meta data on multiple disciplines and at least 1 author from a non-HEI) | QED | Scopus; Dimensions | 3-Behaviour |
| | | To what extent has FLF contributed to new investment into R&I from outside government? | Number (and value) of grants from unique sources outside research councils and IUK since the start of fellowship | TBE/ QED* | Researchfish; career tracker surveys | 2-Learning |
| | | | R&D expenditure | TBE | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; Higher Education Statistics Agency (HESA) data | 3-Behaviour |
|  | To what extent, and how, has FLF delivered highly skilled R&I leaders of the future? | To what extent has FLF developed leadership and research skills and capabilities of fellows? | Number of PhDs supervised in past year | TBE | Career tracker surveys | 3-Behaviour |
| | | | Number of research staff managed (excluding PhD students) | TBE | Career tracker surveys | 3-Behaviour |

| | | | | Support provided to other members of team and contribution to their grant applications etc. | TBE | Career tracker surveys | 3-Behaviour |
|---|---|---|---|---|---|---|---|
| | | | | New research methodologies/skills developed | TBE | Career tracker surveys | 3-Behaviour |
| | | | | Advancements in band/grade (academia) or salary/title (business) | TBE | Career tracker surveys | 4-Impact |
| | | | | Receipt of awards and recognition | TBE | Researchfish; Career tracker surveys | 2-Learning |
| | | | | Perspectives on what kind of leadership has been enabled through FLF with illustrative examples (probing for thought leadership) | TBE | Career tracker surveys and key informant interviews | 4-Impact |
| | | | To what extent has FLF increased careers in R&I within new and novel areas? | New teams/research groups established since start of fellowship | TBE | Key informant interviews; career tracker surveys; host survey | 3-Behaviour |
| | | | | Number of new startups/spin offs | TBE | Researchfish; career tracker surveys; host survey, HESA Higher Education Business and Community Interaction (HE-BCI) data, FAME data, Business Structure Database (BSD) data | 4-Impact |
| | | | | Number of patents filed | QED | Dimensions; Researchfish; career tracker surveys; Espacenet | 3-Behaviour |
| | | | | Researchers who have pursued careers outside academia | TBE | Career tracker surveys | 3-Behaviour |

| | | | New research methodologies/skills developed | TBE | Career tracker surveys | 3-Behaviour |
|---|---|---|---|---|---|---|
| | | | New work generated in business settings | TBE | Career tracker surveys; host survey | 4-Behaviour |
| | | To what extent has FLF developed the innovation and commercialisation skills and capacities of fellows? | Number of patents filed | QED | Dimensions; Researchfish; career tracker surveys; Espacenet | 3-Behaviour |
| | | | Number of patents granted | QED | Dimensions; Researchfish; career tracker surveys; Espacenet | 4-Impact |
| | | | Any new products, services or processes realised | TBE | Researchfish; career tracker surveys | 3-Behaviour |
| | | | Number of new startups/spin offs | TBE | Researchfish; career tracker survey; host survey; HE-BCI data; FAME data; BSD data; Beauhurst data | 4-Impact |
| | | | Perspectives and examples of how FLF has supported innovation intensity of the host/fellows, partnership development, revenue generation of hosts and commercialisation readiness in fellows | TBE | Key informant interviews | 4-Impact |
| | | | Any new products, services or processes realised | TBE | Researchfish; career tracker surveys | 3-Behaviour |
| | | | Number of new startups/spin offs | TBE | Researchfish; career tracker survey; host survey; HE-BCI data; FAME data; BSD data; Beauhurst data | 4-Impact |
| | | | Perspectives and examples of how FLF has supported innovation | TBE | Key informant interviews | 4-Impact |

| | | | intensity of the host/fellows, partnership development, revenue generation of hosts and commercialisation readiness in fellows | | | |
|---|---|---|---|---|---|---|
| | To what extent, and how, does FLF make the UK an attractive place for future R&I leaders? | To what extent has FLF developed, attracted and retained talent (fellows and associated teams) to the UK? | Changes in contract type (e.g. open ended, fixed term) during course of fellowship | TBE | Career tracker surveys | 3-Behaviour |
| | | | Moving location (beyond UK) | TBE | Career tracker surveys; Scopus; Dimensions; desk research (e.g. LinkedIn profiles) | 3-Behaviour |
| | | | Proportion of fellows still in research at a given point in time | TBE | Career tracker surveys | 4-Impact |
| | | | Promotion of non-UK staff | TBE | Host survey; key informant interviews | 3-Behaviour |
| | | | Perceptions on the extent to which FLF has influenced the reputation of the UK as a place to pursue a career in research or innovation, and if so how | TBE | Key informant interviews; host survey; career tracker surveys | 3-Behaviour |
| | How has FLF led to a change in behaviour for early career researchers, innovators and hosts? | To what extent has FLF developed a more equal, diverse and inclusive R&I workforce? | Perception of diversity in teams/host departments + TBC for other EDI measures based on engagement with NEDIAL | TBE | Key informant interviews | 3-Behaviour |
| | | | Diversity characteristics of fellows and the research teams working with fellows (e.g. gender, ethnicity, career stage) | TBE | Career tracker surveys; UK management information | 3-Behaviour |
| | | | Promotion of staff from ethnic minority groups | TBE | Host survey; key informant interviews | 3-Behaviour |
| | | | Promotion of non-UK staff | TBE | Host survey; key informant interviews | 3-Behaviour |

| | | | Promotion of female staff | TBE | Host survey; key informant interviews | 3-Behaviour |
|---|---|---|---|---|---|---|
| | | | Promotion of disabled staff | TBE | Host survey; key informant interviews | 3-Behaviour |
| | | To what extent, and how, has FLF influenced risk taking by fellows, hosts and wider teams in novel areas of research? | R&D expenditure | TBE | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; HESA data | 3-Behaviour |
| | | | Number of new startups/spin offs | TBE | Researchfish; career tracker surveys; host survey; HE-BCI data; FAME data; BSD data; Beauhurst data | 4-Impact |
| | | | Perception of risk taking and ability to do so by fellows and teams | TBE | Key informant interviews | 3-Behaviour |
| | | To what extent have host organisations promoted and supported the FLF scheme and delivered against expectations / commitments for research or innovation support? | Staff satisfaction, engagement, and commitment | TBE | Career tracker surveys; host survey; key informant interviews | 3-Behaviour |
| | | | R&D expenditure | TBE | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; HESA data | 3-Behaviour |
| | | | Staff training and development expense per annum on leadership and commercialisation/innovation training | TBE | Host survey; key informant interviews | 3-Behaviour |
| | | | Additional support provided to fellows (both by central FLF and by hosts) | TBE | Key informant interviews; host survey | 3-Behaviour |
| | | | New teams/research groups established since start of fellowship | TBE | Key informant interviews; career tracker surveys; host survey | 3-Behaviour |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | Examples of where FLF has set precedents or influenced host organisations' policies and values around support for R&I, EDI or UKRI policy goals | TBE | Key informant interviews and host survey | 3-Behaviour |
|  | To what extent and how has the FLF delivered wider knowledge, economic and societal impacts? | What has been the wider, overall impact of the FLF on UK R&I expertise and on other parts of UKRI practice? | Examples of wider, overall impact of the FLF on UK R&I | TBE | Key informant interviews | 4-Impact |
| | | What has been the wider, overall economic impact of the FLF, including the economic value of non-market impacts? | TBC | - | - | 4- Impact |
| | | What has been the wider, overall societal impact of the FLF? | Examples of wider, overall societal impact of the FLF | | Key informant interviews | 4-Impact |
| | Based on the overall, estimated impact of the FLF, to what extent does the FLF represent value for money? | To what extent does the FLF represent value for money in absolute terms? | TBC | - | - | 5-ROI |

| | | | TBC | - | - | 5-ROI |
|---|---|---|---|---|---|---|
| | | To what extent does the FLF represent value for money compared to other possible alternative ways of achieving the same impacts? | | | | |

*Use as QED or TBE metric to be confirmed pending availability of Researchfish and UKRI management information data on unsuccessful applicants to FLF who are in receipt of other UKRI grants.

### 5.2.1. Impact evaluation analytical approaches

The impact evaluation will be analysed through multiple lenses and analytical approaches that are further elaborated on below.

## Kirkpatrick model analysis

As listed in Table 5 and Table 6, the evaluation metrics for both process and impact have been categorised against what we are terming as Kirkpatrick's 'hierarchy of impact', denoted by L1-L5 where 5 represents the highest level of impact achieved in the longer term with clear and quantifiable returns on investment evident. As evidence against the metrics captured during the evaluation, the evidence will be grouped and assessed collectively based on the Kirkpatrick levels to determine where the programme sits in the hierarchy. We will explore whether the outputs of this assessment can be presented visually by showing the extent to which L1 has been achieved, and L2, and so on. As there might be levels where evidence is sparse or only a few outputs, outcomes or impacts have emerged at the point in time when the evaluation takes place.

## QED counterfactual analysis

For the QED analysis, once a balanced counterfactual group is generated, the average treatment effect on the treated will be estimated using traditional parametric analysis techniques (for example, traditional regression techniques such as OLS or ANCOVA). The analysis will primarily focus on those outcomes identified as relevant to the QED in Table 6. The sensitivity[37] of the analysis to the choice of counterfactual group will be investigated by varying the following : a) the model estimating the propensity score, b) the closeness of the match between treated and unsuccessful comparison, and c) the pool of potential individuals for matching. Under the latter sensitivity analysis, the effects of both restricting the pool for matching to those who were invited to interview and broadening the pool to all early career researchers should be considered. Given the macroeconomic shocks that occurred concurrently with the FLF recruitment, experimentation with cohort stratification and the inclusion of cohort controls in the regression analysis should be included as a sensitivity analysis.

The counterfactual impact evaluation will focus on measuring the treatment effect at the individual-level, as wider effects of FLF on institutions, collaborators, and the R&I environment are likely to emerge over a longer time horizon and are harder to accurately measure. Even if the wider effects are observable at the point of analysis, it would still be inappropriate for QED analysis- given that the treatment and counterfactual generation are both at the individual level, it is unlikely the necessary assumptions hold at the institutional level or beyond.

## Context-mechanism-outcome (CMO) analysis

To fortify the QED approach, which will be focused on providing 'treatment effects' by quantifying the evidence of what has been achieved by the fellows versus the counterfactual group, the CMO analysis (a realist evaluation approach) will focus on understanding how and why certain outcomes have been realised. This will provide the programme with a deeper understanding of the interaction between the FLF intervention and the wider R&I environment, providing key learnings for FLF and other fellowship

---

[37] Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system can be divided and allocated to different sources of uncertainty in its inputs

programmes on how impact can be maximised and challenges mitigated. An illustrative example of a CMO analysis output is highlighted in Figure 9 which could be surfaced through key informant interviews and career tracker surveys.

Only a select group of outcomes will be suitable for this analysis as outcomes need to be specific and targeted, and the analysis is more suited to focusing on behaviour related outcomes. Outcomes such as those focused on bringing about societal change or large economic benefits are too broad to be able to elicit meaningful information on context and mechanisms as there are likely to be a large set of these underpinning broad outcomes. Moreover, due to the volume of outcomes being considered in this evaluation, a focused approach is needed to conduct CMO deep dives to keep the tasks manageable and meaningful and to reduce the burden on evaluation participants.

**Figure 9: Illustrative example of CMO analysis**
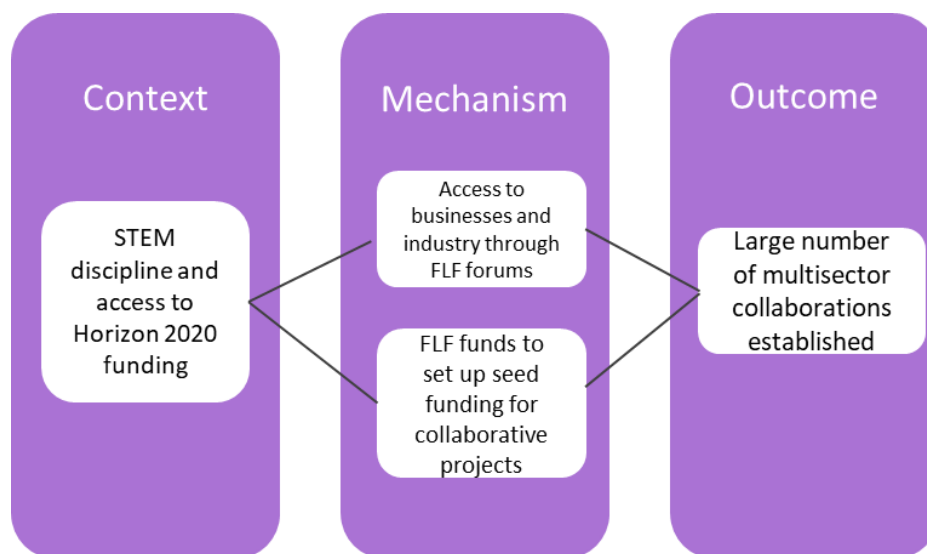


A list of proposed outcomes for this analysis, which have been derived from the evaluation questions from Table 6, are listed below, which are subject to refinement and agreement with UKRI. It might be worthwhile conducting a prioritisation exercise to determine which outcomes would benefit most from a deep CMO assessment.

- FLF has increased multidisciplinary and cross-sector working

- FLF has developed the leadership and research skills of fellows including commercialisation capabilities

- FLF has developed a more equal and diverse R&I workforce.

### 5.2.2. Impact evaluation data collection methods

As indicated in Table 6, the impact evaluation will draw on a wide range of sources, comprising internal UKRI monitoring data and reporting, wider secondary data, career tracker surveys, a host survey and key informant interviews. The data collection methods for the impact evaluation are described further below.

## Internal UKRI monitoring data and reporting

The analysis of internal UKRI monitoring data and reporting, primarily Researchfish data and FLF management information, will be undertaken to surface evidence of activities and progress of awards towards the delivery of FLF outputs, outcomes and impacts. We will also review data within FLF baseline surveys to inform our understanding of the position of fellows prior to receipt of FLF awards. In addition, we will also look to explore whether selected datasets such as Researchfish or management information are available for the counterfactual group of researchers where those researchers are in receipt of other (non-FLF) UKRI grants (relevant metrics are denoted by * in Table 6).

## Secondary data

A range of secondary data sources will also be consulted and, where possible, data extracted and analysed. During the development of this evaluation framework, we have identified which secondary data sources could have potential utility for our impact evaluation (see Table 6). Importantly, incorporation of data from these secondary data sources will depend both on the ability to match FLF beneficiaries (i.e. fellows or businesses) within the databases and, where sources are fee-based, on the provision of access through UKRI's existing subscriptions/licenses.

- **Dimensions (including Altmetric):** The Dimensions database would be used to collect data on Field-Weighted Citation Impact and collaborative publications of FLF fellows and those within the QED) counterfactual group. The Altmetric database would be used to collect data on Attention Scores of FLF fellows and those within the counterfactual group.

- **Scopus:** The Scopus database would be used to collect data on Field-Weighted Citation Impact and collaborative publications of FLF fellows and those within the QED) counterfactual group.

- **FAME data:** The FAME database contains data on private company information in the UK and Ireland. FAME represents a potential source of data on R&D expenditure in business hosts, and potentially of startups/spin offs associated with FLF fellows.

- **BSD data:** The Office for National Statistics Business Structure Database (BSD) contains data on UK companies compiled through the Interdepartmental Business Register. Like FAME, BSD represents a potential source of data R&D expenditure in business hosts, and potentially of startups/spin offs associated with FLF fellows.

- **Beauhurst data:** The Beauhurst database contains data on high-growth companies in the UK. Beauhurst also represents a potential source of data on R&D expenditure in business hosts, and potentially of startups/spin offs associated with FLF fellows.

- **Espacenet data:** The Espacenet database contains data on patent documents and publications. Espacenet represents a potential source of data on patent applications and licenses of FLF fellows.

- **HESA, including HE-BCI data:** The HESA database contains data on various aspects of the UK higher education sector. The HESA HE BCI survey contains data on HE institutions' social, community and cultural engagement, intellectual property, startups and spin-outs. HESA and HE-BCI present a potential source of data on startups/spin offs associated with FLF fellows.

- **LinkedIn data:** LinkedIn professional profiles represent a potential data source on the location of FLF fellows.

## Career tracker surveys

To capture data on more qualitative programme outputs, outcomes and impacts, career tracker surveys will be implemented across all FLF fellows. Survey questions will be based on the evaluation framework and based on stakeholder type (e.g. award date, discipline) with a set of core questions answered by all fellows. The questions will seek to understand both the nature of the outputs, outcomes and impacts achieved by fellows, as well as the processes through which these have been achieved. The evaluation team will design the surveys to ensure questions focus on surfacing information not already captured by fellows' existing reporting requirements, i.e. Researchfish submissions. The career tracker surveys are planned to be conducted at the start of year 2 and year 3 of the evaluation (see Chapter 6), though timing will be finalised in consultation with UKRI for timely data collection. The surveys will be implemented using an online survey platform, e.g. SmartSurvey.

## Host survey

Alongside the career tracker surveys, we will also conduct a one-off host institution survey. The aim of the host survey will be to capture further data on the forms of support provided to FLF fellows by hosts (in addition to that gleaned from career tracker surveys), and on the impact of fellowships on host institutions' processes and culture. As with the career tracker surveys, the host survey will be designed with a view to minimise duplication of other data collection tools. The planned timing of the host survey is at the start of year 3 of the evaluation, to coincide with the second career tracker survey. The survey will also be implemented using an online survey platform.

## Key informant interviews

To further strengthen our understanding of the nature and pathways of outputs, outcomes and impacts of FLF fellows, we will conduct a programme of interviews. Rather than seeking to capture outputs, outcomes and impacts (this being primarily identified through other impact evaluation data collection tasks, e.g. career tracker and host surveys) a key aim of the interviews will be to delve into the processes through which outputs, outcomes and impacts have been achieved. The impact evaluation interviews will engage:

- Up to 20 FLF fellows (at least three per round of funding)

- Up to eight members of staff at host institutions of FLF fellows

- Up to five interviews with wider sector bodies/leaders.

- Up to four fellows other schemes

As with the process evaluation interviews, interviewees will be selected using a stratified, purposive sampling approach involving the stratification of fellows within each round by defined criteria and the

purposive selection of candidates to ensure a mix within each cohort and across the sample as a whole. In selecting fellows and host institutions for inclusion in the sample, we will seek to identify a new set of fellows to those engaged in the process evaluation interviews, thereby capturing fresh perspectives and reducing the burden on individuals. As part of our sampling for fellows, we will seek to identify a number of FLF fellows who have also received funding from other fellowship schemes. This will support probing of the value added of FLF compared to other schemes through interview questions. The eight host institutions will also be selected to ensure a mix of institution types, both in terms of sector (business/academia) and size. We will also conduct up to five impact evaluation interviews with sector bodies/leaders, focusing on key leaders of other early career fellowship programmes, to understand their perspective on the impact of FLF, and how this compares to other programmes. All interviews will follow semi-structured format using a protocol designed around the evaluation framework and Levels 2-4 of the Kirkpatrick model.

# 6. Plan and deliverables

This chapter sets out our plan for implementation of the evaluation, including the main evaluation deliverables and deadlines.

## 6.1. Evaluation plan

A Gantt chart visualising the planned timeframe for the implementation of evaluation activities is presented in Figure 10 below.

**Figure 10: FLF evaluation Gantt**

| Task | Year 1: 22-23 | | | | Year 2: 23-24 | | | | Year 3: 24-25 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jun.22 | Sep.22 | Dec.22 | Mar.23 | Jun.23 | Sep.23 | Dec.23 | Mar.24 | Jun.24 | Sep.24 | Dec.24 | Jun.25 |
| **Phase 1: Evaluation framework** | | | | | | | | | | | | |
| Documentary analysis | | | | | | | | | | | | |
| Evaluation framework report | | | | | | | | | | | | |
| **Phase 2: Process evaluation** | | | | | | | | | | | | |
| *Data collection* | | | | | | | | | | | | |
| Documentary analysis and literature review | | | | | | | | | | | | |
| Management information analysis | | | | | | | | | | | | |
| Key informant interviews | | | | | | | | | | | | |
| *Data analysis* | | | | | | | | | | | | |
| Process mapping | | | | | | | | | | | | |
| Cause consequence analysis | | | | | | | | | | | | |
| Case study development | | | | | | | | | | | | |
| Comparator case studies development | | | | | | | | | | | | |
| *Reporting* | | | | | | | | | | | | |
| Process evaluation report | | | | | | | | | | | | |
| **Phase 3: Interim impact evaluation and economic feasibility report** | | | | | | | | | | | | |
| *Data collection* | | | | | | | | | | | | |
| Internal UKRI montoring and data | | | | | | | | | | | | |
| Secondary data collection | | | | | | | | | | | | |
| First career tracker and host survey | | | | | | | | | | | | |
| *Data analysis* | | | | | | | | | | | | |
| QED counterfactual analysis | | | | | | | | | | | | |
| Economic methods feasibility assessment | | | | | | | | | | | | |
| Economic scoping and feasiblity workshop | | | | | | | | | | | | |
| *Reporting* | | | | | | | | | | | | |
| Interim impact evaluation and economic feasibility report | | | | | | | | | | | | |
| **Phase 4: Final impact evaluation report and economic evaluation** | | | | | | | | | | | | |
| *Data collection* | | | | | | | | | | | | |
| Second career tracker survey | | | | | | | | | | | | |
| Key Informant Interviews | | | | | | | | | | | | |
| Economic datasets gathering | | | | | | | | | | | | |
| *Data analysis* | | | | | | | | | | | | |
| Context-mechanism-outcome analysis | | | | | | | | | | | | |
| Kirkpatrick model analysis | | | | | | | | | | | | |
| Monetisation and Modelling Analysis | | | | | | | | | | | | |
| *Reporting* | | | | | | | | | | | | |
| Final impact evaluation report and economic evaluation report | | | | | | | | | | | | |
| **Project management and quality assurance** | | | | | | | | | | | | |

Overall task duration
Overall sub-task duration
Sub-task duration
Report due

## 6.2.   Evaluation deliverables and deadlines

Table 7 shows the main deliverables and their associated deadlines for the project.

Table 7: Main deliverables and deadlines

| Deliverables | Phase | Deadline |
|---|---|---|
| Fortnightly/monthly catch-ups with the UKRI evaluation points of contact with updates on progress and to revise the evaluation approach as needed | All | Throughout project |
| Quarterly presentations to the NPIF Evaluation Oversight Board, FLF Evaluation Advisory Board and FLF Board | | Throughout project |
| Kick-off meeting with UKRI | 1 | Jun-22 |
| Evaluation framework report | 1 | Aug-22 |
| Process evaluation report | 2 | Jun-23 |
| Economic scoping and feasibility workshop | 3 | Jan-24 |
| Interim impact evaluation report and economic evaluation | | Jun-24 |
| Final impact evaluation report and economic evaluation report | 4 | Jun-25 |

# 7. Evaluation limitations and risks

This chapter considers key limitations and potential risks to our evaluation of the FLF.

## 7.1. Limitations of the evaluation approach

The approach and methodologies outlined in the chapters above will provide a wide-ranging set of data and evidence around the causal impact of the FLF, and how these impacts were achieved. However, our approach to the evaluation is also subject to a number of important limitations.

First, exogenous shocks, such as Covid-19, have had uneven impacts on FLF cohorts and disciplines, presenting challenges to fair evaluation. Earlier cohorts have been particularly disadvantaged by the pandemic and related restrictions, for example. While we account for this in the QED approach—and set out Covid-19 as an important contextual factor in the TBE strand of the impact evaluation—it may be difficult to make fair appraisals and generate rigorous cross-cohort comparisons. This will have no impact on our ability to make causal claims around the FLF as a whole, given that unsuccessful applicants will have been similarly impacted by Covid-19 and restrictions.

Second, while our approach aims to be comprehensive and cover differential impacts of the FLF on various sub-groups, given the small sample size of the business awardees and AHSS awardees, as highlighted in Chapter 4, a nuanced assessment of the impact on these sub-groups will likely be lost in the QED approach. To some extent this nuance can be accounted for with the TBE methods, as we can conduct interviews and gather qualitative evidence from business and AHSS awardees.

Third, given that much of the anticipated impact of the FLF will only emerge over a lengthy time horizon, the evaluation will not be able to capture its long-term impact in full. Ideally, the evaluation would involve a long-term follow up and assessment of the fellows and unsuccessful applicants to track the longer-term impacts of the programme. What we are proposing is a step in this direction, setting out a range of indicators that can be used to assess whether the programme is *on track* to achieve longer-term desired outcomes and impacts. Additionally, the evaluation has been designed in a way to ensure it is repeatable and that longer-term impacts could be captured in full at a later date using the same methods and framework.

Fourth, the lack of a comprehensive baseline analysis of the state of R&I fellowships and culture in the UK could mean that attributing large scale impacts to the FLF is challenging. If the strengths and weaknesses of the current R&I landscape are not well understood, it could be difficult to understand how the FLF has improved upon them, if at all, and what impact this has had. However, RAND already has a specialist understanding of the R&I landscape in the UK and, specifically, the challenges around early career

researchers. Additionally, the review of baseline fellow surveys during the impact assessment can bolster this understanding and put the accomplishments of the FLF in context.

Finally, as the qualitative intended impacts of the FLF, such as leadership and improvements in culture, are difficult to capture directly, the evaluation is dependent on proxy measures. For example, to understand the impact of the FLF on leadership skills, we have included indicators such as 'number of research staff (that fellows have) managed'. We have circumvented this through using a wide variety of indicators and are also conducting a number of interviews with diverse stakeholders to capture qualitative perceptions of the impact of the FLF on those softer skills.

It is also worth noting that in our evaluation methodology, whilst we have determined the categorisation of a metric as either QED or TBE based on our assessment of data availability at the time of submitting this evaluation framework report, there is a possibility that the collection of data for some QED metrics may encounter challenges during implementation, in which case those metrics may be dropped, revised or incorporated into the broader TBE approach. This is a limitation linked to data availability and access.

## 7.2.  Evaluation risks and mitigations

While the previous section set out broad limitations of the evaluation, Table 8 below presents a more specific set of pertinent project risks and how we propose to mitigate them.

**Table 8: Evaluation risks and mitigations**

| | |
|---|---|
| **Risk:** QED approach is unfeasible<br>**Likelihood:** Low<br>**Impact:** High | If a QED approach is impractical because of the lack of a suitable comparator group or other reason, we will instead adopt a realist approach and aim to understand the 'value added/additionality' of the intervention. So, instead of focussing on the individual fellow as the unit of analysis, we would look at the wider, aggregate impacts of the programme as a whole. This would be primarily done in a qualitative way through interviews, focus groups and/or surveys with key policy and business stakeholders and successful fellows as well as non FLF researchers, where feasible and appropriate. |
| **Risk:** Impacts cannot be fully attributed to the FLF<br>**Likelihood:** Medium<br>**Impact:** High | Attributing impacts directly to FLF can be a challenge given that it is occurring in a complex policy environment with its intended outcomes having several underlying causes. Adopting a hybrid TBE and QED approach can resolve this issue. The QED approach can understand the causal impact the fund has made on fellows through comparison with unsuccessful applicants, while the TBE approach is useful in understanding the broader macroeconomic impacts of the fund. However, understanding and attributing spillovers to the FLF, especially in a quantitative sense, will still be challenging given timelines of the evaluation but a qualitative assessment of interim measures might be more useful. |
| **Risk:** Impacts are too diverse to fully capture<br>**Likelihood:** Low<br>**Impact:** Medium | FLF has a wide range of ambitions and anticipated impacts, and there is a risk that we will not be able to identify relevant outcome measures and proxy measures that fully reflect the broad scope of FLF. By combining QED with a realist TBE approach, our impact evaluation will ensure a mix of quantitative output-focussed measures with a wider range of qualitative and proxy measures intended to capture FLF's wider and 'softer' impacts. |

| | |
|---|---|
| **Risk:** Impacts have not fully emerged during the timeframe of the evaluation<br><br>Likelihood: High<br><br>Impact: Medium | FLF is a large, ambitious programme, whose intended impacts will take several years to emerge in full. Given that the evaluation will conclude in 2025 at the very latest, there is a risk that these impacts will not be fully fledged and captured. However, by rigorously assessing outputs and outcomes which are apparent, we can determine whether the FLF is reasonably *on track* to accomplish its more longer-term objectives and what further actions are required to maximise impact. Additionally, our evaluation design is fully repeatable so that, several years later, the evaluation could be repeated to track those longer-term impacts. |
| **Risk:** Low engagement of stakeholders in surveys and interviews<br><br>Likelihood: High<br><br>Impact: Medium | While engagement with FLF stakeholders through surveys and interviews is an important data source for the evaluation, there is a risk of a low response rate, especially among unsuccessful applicants. To mitigate this, we would value support from UKRI in distributing the survey and in engaging interviewees. We will track survey coverage against set targets and take actions to achieve sufficient samples for robust analysis. To ensure that unsuccessful applicants are surveyed, we can provide monetary incentives. |
| **Risk:** Secondary datasets cannot be accessed or do not contain suitable data<br><br>Likelihood: Medium<br><br>Impact: Medium | As part of our impact evaluation, we will consult a range of secondary data sources to determine whether data can be collected to inform our evaluation. However, incorporation of data from these sources will depend both on the ability to match FLF beneficiaries (e.g. fellows or businesses) within these databases and, where sources are fee-based, on the provision of access through UKRI's existing subscriptions/licenses. Where we cannot obtain access to data, or the data available is not suitable (e.g. cannot be effectively matched), we will need to rely on other data sources, including in some cases qualitative data sources such as interviews, to assess the performance of FLF against a given evaluation metric. |
| **Risk:** Key personnel leave the evaluation team<br><br>Likelihood: Medium<br><br>Impact: Low | RAND is a large, well-staffed research organisation with over 100 experienced researchers, so staff turnover does not represent a major risk to the evaluation. If the project leader, Dr Sana Zakaria, were unavailable then Research Group Director, Dr Susan Guthrie would assume the role. If the project manager, Dr Joe Francombe, were unavailable then Dr Dadiso Motsi-Omoijiade would take over. |
| **Risk:** Stakeholders' recall is limited<br><br>Likelihood: Low<br><br>Impact: Medium | The TBE strand of the evaluation is dependent on the qualitative, subjective views of stakeholders, whose memory may have faded over time. Subsequently, there is a risk that key details are lost or mistakenly recalled, and that the evaluation is less objective. This is a minor risk as we are speaking to a number of individuals, so we are not dependent on any one individuals' memory. Second, given that the FLF is a relatively recent funding programme, recall will likely be strong. |
| **Risk:** Impacts are over-estimated due to dependence on self-reported data (e.g. Researchfish)<br><br>Likelihood: High<br><br>Impact: Low | While some data sources may overestimate the impact of the FLF, we are using a wide variety of secondary data sources which can mitigate the limitations of individual datasets. Triangulating methods and sources across the evaluation can provide a robust picture of the outcomes and impacts of the FLF. |

# Annex A. Review of existing FLF audits, reporting and reviews

This annex contains a more detailed description of the steps taken in the review of existing FLF audits, reports and reviews, and describes how the review of these documents fed into the process evaluation framework.

**Step 1: Identifying relevant FLF documents**

Documents deemed relevant and significant to informing the process evaluation framework questions were identified in consultation with the UKRI. These were a) the Government Internal Audit Agency (GIAA) UKRI FLF Audit, which aimed to assess the effectiveness of the organisation, structure and processes in place to ensure that the high level aims of the scheme are met and that risks are managed; b) the BEIS Critical Friend review of the FLF which provided a snapshot review of the FLF reflecting the conclusions of an independent Assurance Review Team; c) BEIS FLF quarterly reports – 11 reports reviewed from Q2-Q4 2019-20, Q1 -Q4 2020-21 and Q1-Q4 2021-22, based on RAG (Red-Amber-Green) ratings covering finance, people, milestones and benefits metrics.

**Step 2: Document review and extraction template**

After identifying the key documents, all the relevant literature was reviewed, and pertinent information was documented using an extraction template based on the initial process evaluation framework questions. The extraction template was used to pull out and cluster sections in the reviewed documents related to each process evaluation question. This was done in order to identify a) the extent to which the information and data in the reviewed documents addressed the process evaluation questions and, b) any process-related considerations that were not addressed in the process evaluation questions. The extraction template further analysed each of the FLF quarterly reports sequentially (from 2019 to 2022) to get a clearer view of the evolution of FLF RAG ratings and the FLF processes over time. This approach to the literature and document review facilitated the identification of key process-related themes and trends (Step 3) and informed the further revision and refinement of the process evaluation framework questions (Step 4).

**Step 3: Identification of trends and themes**

Several key trends and themes relevant to the process evaluation framework were identified during the document review process. Of note were the inter-connected observations to do with the effects of Covid-19, resourcing and staffing, and governance.

*a) Effects of Covid-19*

The Covid-19 pandemic had a significant impact on FLF processes from the first reviewed quarterly report Q2 2019 till the latest report from Q4 2021-22. These effects ranged from delays, extensions, unmet targets, and forecasting and staffing challenges; to lockdowns affecting the quality, range and diversity of applicants. The Q2 2019 report[38] highlighted how the ministerial announcement around FLF was delayed as Covid-19 announcements took priority. Similarly, the Q4 2020 report[39] explained how the decision to extend the closing date for the round five competition by one month was taken as a direct response to Covid-19 and in Q1 2020-21 the timeline for undertaking the assessment of proposals was extended by one month to nine months in total. This resulted in further delays in start time and led to changes to activity profiles leading the FLF programme to submit a re-profile request aimed at re-aligning the programme's forecast.[40] The re-profiling request highlights steps taken to mitigate for the effects of Covid-19 on the FLF, where the pushing back of the spend profile by at least one quarter resulted in finance leads working on slippage estimations.[41] Other challenges brought on by Covid-19 have to do with staff shortages, where the Q1 2020/21 report highlighted how the FLF team was working at reduced capacity (60%) during a time where there was increased demand (through round four and round five) resulting in the need to identify additional delivery resources.[42] Significantly, the Q3 2020/21 report explained how the spend quarter for round 6 was moved by one quarter to accommodate changes and to allow a change in the deadline to full submissions to ensure that the diversity of applications was not impacted due to the differential impact of lockdown.[43] In this way, Covid-19 resulted not only in delays to timelines (which also had knock on financial and forecasting impacts), but also significantly impacted staffing and resourcing. as well as potentially the diversity of FLF programme applicants as alluded to (but not substantiated) by the Q3 2020/21 report.[44]

*b) Resources/Staffing*

The second significant theme identified in the document review has to do with staffing and resourcing. Here, the quarterly reports highlighted various concerns including the number of reviewers, Covid-19 staff reductions and subsequent recruitment drives, and the loaning of FLF staff to other UKRI departments. As has been previously mentioned, FLF teams were working at reduced capacity due to the pandemic[45] and due to increased demand (up by over 50%) particularly through round five of the scheme.[46] Unlike the quarterly reports which made the causal link between Covid-19 and staffing and resourcing challenges, the Critical Friend Review attributed delays exclusively to limited resources (including staffing) in part due to the high levels of good quality applications which required a huge resource to assess using the declared process and criteria, further stating that "the available resource was insufficient to complete the exercise on

---

[38] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Reports – Future Leaders Fellowships Programme," 2019.

[39] Department for Business, Energy, and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2020.

[40] Department for Business, Energy, and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2021.

[41] Department for Business, Energy and Industrial Strategy (BEIS).

[42] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2020.

[43] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2021.

[44] Department for Business, Energy and Industrial Strategy (BEIS).

[45] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2020.

[46] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2021.

time and the decision-making process to address the issue was inadequate."[47] The review found that it was highly likely that resources would be further constrained for the pre-award element of rounds 7 and 8 due to high interest in the scheme resulting in an increased volume of applications above the level seen in round 6, leading to the drafting of recommendations by the operational review aimed at making the process more manageable.[48] The FLF programme showed resilience and adaptability, responding through recruitment where over 1000 additional people were recruited to FLF Peer Review College to support peer review of proposals in Q4 2020/21).[49] However, concerns were raised in the Q2 2021/22 report about the uncertainties being faced by the FLF team as team members began to be 'loaned out' to other parts of UKRI pending the securing of future funding, presenting a risk of losing experienced staff to other UKRI roles that are perceived to be more secure.[50]

c) *Governance*

The final key theme observed has to do with governance. The Critical Friend Review noted the delays in the programme due to staffing and other considerations but was concerned about the governance of the FLF scheme, particularly the decision-making processes to address delays and other issues, which it described as inadequate. Here, the review team noted that "the SRO did not have the necessary delegated authority to engage additional resources, nor the authority to change the process"[51] and explained how there was "no clear route in the governance structure to escalate this issue with the end result that the decision to delay the awards was made after discussion with senior stakeholders across UKRI (including the CE), based on, by that point, a very limited number of options"[52]. The Critical Friend Review raised additional concerns around benefits reporting and management and the need to further clarify roles and decision-making where, for example, it is reportedly unclear what the defined roles, duties, responsibilities and accountabilities for the SRO and the Programme Sponsor are.[53] These concerns around the governance and decision-making processes to do with the FLF were echoed in the GIAA audit which also highlighted concerns around the inadequate coordination of Risk Management and insufficient action tracking.[54]

**Step 4: Refining process evaluation framework questions**

The final step in the document review consisted of assessing and evaluating the validity of the process evaluation questions as set out in the evaluation RFP and sense-checking these based on the scope and nature of evidence provided in the reviewed documents. Whilst the majority of questions were found to be relevant with no revisions necessary, two of the questions were refined and adjusted as detailed in Table 9. In addition to informing the revision of the process evaluation framework, the document review also allowed

---

[47] David Noble, Amy Rogers, and Owen Jackson, "Critical Friend Review Project Title: Future Leaders Fellowships" (Department for Business Energy and Industrial Strategy, October 19, 2021).

[48] Noble, Rogers, and Jackson.

[49] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2021.

[50] Department for Business, Energy and Industrial Strategy (BEIS), "BEIS Quarterly Report – Future Leaders Fellowships Programme," 2022.

[51] Noble, Rogers, and Jackson, "Critical Friend Review Project Title: Future Leaders Fellowships."

[52] Noble, Rogers, and Jackson.

[53] Government Internal Audit Agency, "UK Research and Innovation (UKRI) Future Leaders Fellowship and Innovation Scholars" (Government Internal Audit Agency, February 10, 2022).

[54] Government Internal Audit Agency.

for the refining of the metrics and key considerations relevant to each question as presented in the final iteration of the process evaluation framework.

Table 9: Process evaluation original questions assessment and revisions

| Original process evaluation question (as per evaluation RFP) | Assessment of original question on document review | Revised process evaluation question |
|---|---|---|
| (To what extent) are targets for FLF inputs and outputs being met? | Question lacks the specificity needed for analytical clarity. Reviewed documents highlighted categories and sub-categories of inputs and outputs around finance, awardee and applicant targets. | Q1. To what extent have the FLF financial, awardee and applicant targets been met? |
| To what extent (and why) has the funding scheme hit (or not hit) its target audience, and what may be the consequences and implications of this? | Relevant findings on why the funding scheme did not hit its target audience (mostly Covid-related) did not cover the extent to which the scheme hit its target audience or the consequences/implications. This question was deleted with considerations around the extent to which the scheme hit or did not hit the target audience addressed in question 1 and the consequences and implication of this evidenced in questions 6 and 7. | Q2. To what extent and how has FLF processes supported multidisciplinarity in the fellowship scheme? |
| (To what extent) has FLF stimulated multi- and interdisciplinary research and innovation (MIDRI) fellowships? | Document review found no revision necessary. | Q3. To what extent is the premise of the FLF scheme unique in the sector? |
| (To what extent) are FLF fellowships additional to other schemes supported by UKRI and other UK / international funders? | Document review found no revision necessary. | Q4. To what extent is the premise of the FLF scheme unique in the sector? |
| (To what extent) has FLF delivered effective post award management to support the professional development of the fellows? | Document review found no revision necessary. | Q5. To what extent has FLF delivered effective post award management to support the professional development of the fellows? |
| What, in practice, is felt to be working more / less well regarding the delivery of the fund, and why? | Document review found no revision necessary. | Q6. What, in practice, is felt to be working more / less well regarding the delivery of the fund? |
| What are the unexpected barriers or facilitators to the FLF processes and the delivery of the anticipated outcomes, if any? | Document review found no revision necessary. | Q7. What are the unexpected barriers or facilitators to the FLF processes and the delivery of the anticipated outcomes, if any? |
| What lessons are there for future rounds / similar schemes? | Document review found no revision necessary. | Q8. What lessons are there for future rounds / similar schemes? |

Source: RAND Europe analysis

# Annex B. Detailed impact evaluation framework

This annex presents a detailed version of the impact evaluation framework presented in Chapter 5. The detailed format includes information, for each evaluation metric, on the unit at which data collection is planned to be conducted, as well as the sector and discipline to which the metric applies. With regard to disciplines, Table 10 uses the high-level distinction between STEM and AHSS primarily to indicate the applicability of QED metrics to these broad disciplinary domains, and in particular to highlight that, for some metrics, application to AHSS disciplines may be irrelevant or misleading. In practice, both the QED and TBE aspects of our evaluation will consider potential differences in outputs, outcomes and impacts at a more granular disciplinary level.

Table 10: Impact evaluation framework (detailed format)

| Overarching EQ | Sub EQ | Metric | QED/TBE | Unit (fellow, host or both) | Sector (academia, business or both) | Discipline (STEM, AHSS or all) | Proposed data sources |
|---|---|---|---|---|---|---|---|
| How has FLF shaped the R&I landscape? | To what extent and how has FLF increased high quality and impactful R&I? | Average field-weighted citation impact of researcher | QED | Fellow | Academia | All | Dimensions; Scopus |
| | | Altmetric attention score | QED | Fellow | Academia | All | Altmetric |
| | | Number/nature of and ability to develop grey literature outputs since start of fellowship | TBE | Fellow | Both | All | Researchfish; career tracker surveys |

| | | Contribution to policy making (including participation in national consultations) | QED/TBE* | Fellow | Both | All | Researchfish; career tracker surveys |
|---|---|---|---|---|---|---|---|
| | To what extent and how does the overall FLF fund support wider government objectives? | Perspectives on the extent to which FLF has supported wider government objectives, and if so, how | TBE | Both | Both | All | Key informant interviews |
| | | Examples of FLF supporting wider government objectives | TBE | Both | Both | All | Key informant interviews |
| | | New and existing collaborations (public/third/academic/corporate) | TBE | Fellow | Both | All | Career tracker surveys; Researchfish |
| | | Number of grants from more than one research council since the start of fellowship | QED/TBE* | Fellow | Both | All | UKRI central data; career tracker surveys |
| | To what extent and how has FLF increased MIDRI and cross-sector working, for the fellow, fellow's team and/or collaborators? | Extent of collaborative publications (assessing meta data on multiple disciplines and at least 1 author from a non-HEI) | QED | Fellow | Both | All | Scopus; Dimensions |
| | | Contribution to policy making (including participation in national consultations) | TBE | Fellow | Both | All | Researchfish; career tracker surveys |
| | | Perspectives on the extent to which FLF has increased MIDRI and cross-sector working, and if so, how | TBE | Both | Both | All | Key informant interviews |
| | | Examples of FLF supporting MIDRI and cross-sector working | TBE | Both | Both | All | Key informant interviews |
| | To what extent has FLF increased engagement between industry and | New and existing collaborations (public/third/academic/corporate) | TBE | Fellow | Both | All | Career tracker surveys; Researchfish |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | academia on R&I activities? | Extent of collaborative publications (meta data on multiple disciplines and at least 1 author from a non-HEI) | QED | Fellow | Both | All | Scopus; Dimensions |
| | To what extent has FLF contributed to new investment into R&I from outside government? | Number (and value) of grants from unique sources outside research councils and IUK since the start of fellowship | TBE | Fellow | Both | All | Researchfish; career tracker surveys |
| | | R&D expenditure | TBE | Host | Both | All | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; Higher Education Statistics Agency (HESA) data |
| To what extent, and how, has FLF delivered highly skilled R&I leaders of the future? | To what extent has FLF developed leadership and research skills and capabilities of fellows? | Number of PhDs supervised in past year | TBE | Fellow | Academia | All | Career tracker surveys |
| | | Number of research staff managed (excluding PhD students) | TBE | Fellow | Academia | All | Career tracker surveys |
| | | Support provided to other members of team and contribution to their grant applications etc. | TBE | Fellow | Academia | All | Career tracker surveys |
| | | New research methodologies/skills developed | TBE | Fellow | Both | All | Career tracker surveys |
| | | Advancements in band/grade (academia) or salary/title (business) | TBE | Fellow | Both | All | Career tracker surveys |
| | | Receipt of awards and recognition | TBE | Fellow | Both | All | Researchfish; Career tracker surveys |
| | | Perspectives on what kind of leadership has been enabled through FLF with illustrative examples (probing for thought leadership) | TBE | Fellow | Both | All | Career tracker surveys and key informant interviews |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | New teams/research groups established since start of fellowship | TBE | Both | Both | All | Key informant interviews; career tracker surveys; host survey |
| | | Number of new startups/spin offs | TBE | Both | Both | All | Researchfish; career tracker surveys; host survey, HESA Higher Education Business and Community Interaction (HE-BCI) data, FAME data, Business Structure Database (BSD) data |
| | To what extent has FLF increased careers in R&I within new and novel areas? | Number of patents filed | QED | Fellow | Both | STEM | Dimensions; Researchfish; career tracker surveys; Espacenet |
| | | Fellows who have transitioned to pursue careers outside academia | TBE | Fellow | Academia | All | Career tracker surveys |
| | | New research methodologies/skills developed | TBE | Fellow | Both | All | Career tracker surveys |
| | | New work generated in business settings | TBE | Host | Business | All | Host survey |
| | To what extent has FLF developed the innovation and commercialisation skills and capacities of fellows? | Number of patents filed | QED | Fellow | Both | STEM | Dimensions; Researchfish; career tracker surveys; Espacenet |
| | | Number of patents granted | QED | Fellow | Both | STEM | Dimensions; Researchfish; career tracker surveys; Espacenet |
| | | Any new products, services or processes realised | TBE | Fellow | Both | All | Researchfish; career tracker surveys |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Number of new startups/spin offs | TBE | Both | Both | All | Researchfish; career tracker survey; host survey; HE-BCI data; FAME data; BSD data |
| | | Perspectives and examples of how FLF has supported innovation intensity of the host/fellows, partnership development, revenue generation of hosts and commercialisation readiness in fellows | TBE | Fellow | Both | All | Key informant interviews |
| To what extent, and how, does FLF make the UK an attractive place for future R&I leaders? | To what extent has FLF developed, attracted and retained talent (fellows and associated teams) to the UK? | Changes in contract type (e.g. open ended, fixed term) during course of fellowship | TBE | Fellow | Both | All | Career tracker surveys |
| | | Moving location (beyond UK) | TBE | Fellow | Both | All | Career tracker surveys; Scopus; Dimensions; desk research (e.g. LinkedIn profiles) |
| | | Proportion of fellows still in research at a given point in time | TBE | Fellow | Both | All | Career tracker surveys |
| | | Promotion of non-UK staff | TBE | Host | Both | All | Host survey; key informant interviews |
| | | Perceptions on the extent to which FLF has influenced the reputation of the UK as a place to pursue a career in research or innovation, and if so how | TBE | Both | Both | All | Key informant interviews; host survey; career tracker surveys |
| How has FLF led to a change in behaviour for early career researchers, innovators and hosts? | To what extent has FLF developed a more equal, diverse and inclusive R&I workforce? | Perception of diversity in teams/host departments + TBC for other EDI measures based on engagement with NEDIAL | TBE | Both | Both | All | Key informant interviews |
| | | Diversity characteristics of fellows and the research teams working with fellows (e.g. gender, ethnicity, career stage) | TBE | Fellows | Both | All | Career tracker surveys; Management information analysis |

| | | Promotion of staff from ethnic minority groups | TBE | Host | Both | All | Host survey; key informant interviews |
|---|---|---|---|---|---|---|---|
| | | Promotion of non-UK staff | TBE | Host | Both | All | Host survey; key informant interviews |
| | | Promotion of female staff | TBE | Host | Both | All | Host survey; key informant interviews |
| | | Promotion of disabled staff | TBE | Host | Both | All | Host survey; key informant interviews |
| | To what extent, and how, has FLF influenced risk taking by fellows, hosts and wider teams in novel areas of research? | R&D expenditure | TBE | Host | Both | All | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; HESA data |
| | | Number of new startups/spin offs | TBE | Both | Both | All | Researchfish; career tracker surveys; host survey; HE-BCI data; FAME data; BSD data |
| | | Perception of risk taking and ability to do so by fellows and teams | TBE | Fellow | Both | All | Key informant interviews |
| | To what extent have host organisations promoted and supported the FLF scheme and delivered against expectations / commitments for research or innovation support? | Staff satisfaction, engagement and commitment | TBE | Both | Both | All | Career tracker surveys; host survey; key informant interviews |
| | | R&D expenditure | TBE | Host | Both | All | Host survey; key informant interviews; FAME data; BSD data; Beauhurst data; HESA data |
| | | Staff training and development budget expense per annum on leadership and commercialisation/innovation training | TBE | Host | Both | All | Host survey; key informant interviews |

| | | Additional support provided to fellows (both by central FLF and by hosts) | TBE | Host | Both | All | Key informant interviews; host survey |
|---|---|---|---|---|---|---|---|
| | | New teams/research groups established since start of fellowship | TBE | Both | Both | All | Key informant interviews; career tracker surveys; host survey |
| | | Examples of where FLF has set precedents or influenced host organisations' policies and values around support for R&I, EDI or UKRI policy goals | TBE | Host | Both | All | Key informant interviews |
| To what extent and how has the FLF delivered wider knowledge, economic and societal impacts? | What has been the wider, overall impact of the FLF on UK R&I expertise and on other parts of UKRI practice? | Examples of wider, overall impact of the FLF on UK R&I | TBE | Both | Both | All | Key informant interviews |
| | What has been the wider, overall economic impact of the FLF, including the economic value of non-market impacts? | TBC | - | - | - | - | - |
| | What has been the wider, overall societal impact of the FLF? | Examples of wider, overall societal impact of the FLF | TBE | Both | Both | All | Key informant interviews |
| Based on the overall, estimated impact of the FLF, to what extent does | To what extent does the FLF represent value for money in absolute terms? | TBC | - | - | - | - | - |

| the FLF represent value for money? | To what extent does the FLF represent value for money compared to other possible alternative ways of achieving the same impacts? | TBC | - | - | - | - | - |
|---|---|---|---|---|---|---|---|

*Use as QED or TBE metric to be confirmed pending availability of Researchfish and UKRI management information data on unsuccessful applicants to FLF who are in receipt of other UKRI grants.

# Annex C. QED counterfactual generation

This annex gives more technical detail as to how the counterfactual for the QED analysis was generated under PSM.

As discussed in Chapter 4, we must first verify that PSM satisfies the required assumptions: i) *selection on observables* and ii) the *common support* assumption. Given the metrics available and the application process, as discussed above, we can be somewhat confident that the selection on observables assumption is satisfied. The histograms in Figure 11 allow us to examine the validity of the common support assumption. The variable of interest is the propensity score, as this gives the overall likelihood of treatment. If the successful and unsuccessful applicants are distributed across a common support, i.e., we can find a counterfactual individual with a similar likelihood of treatment for all possible propensity scores, then there would be a non-zero number of individuals in each 'bin' in the histogram for both successful and unsuccessful applicants. Note that this does not mean the shape of the distribution should be the same, but that there should be a non-zero number of individuals with each propensity score in both treatment and non-treated groups. We see immediately that this does not hold true at both the highest and lowest likelihoods of treatment, as expected. The *common support* assumption fails, so we will need to restrict our counterfactual to only those who are on the common support.

## Figure 11: Sift score supports by round sub-groups and application outcome
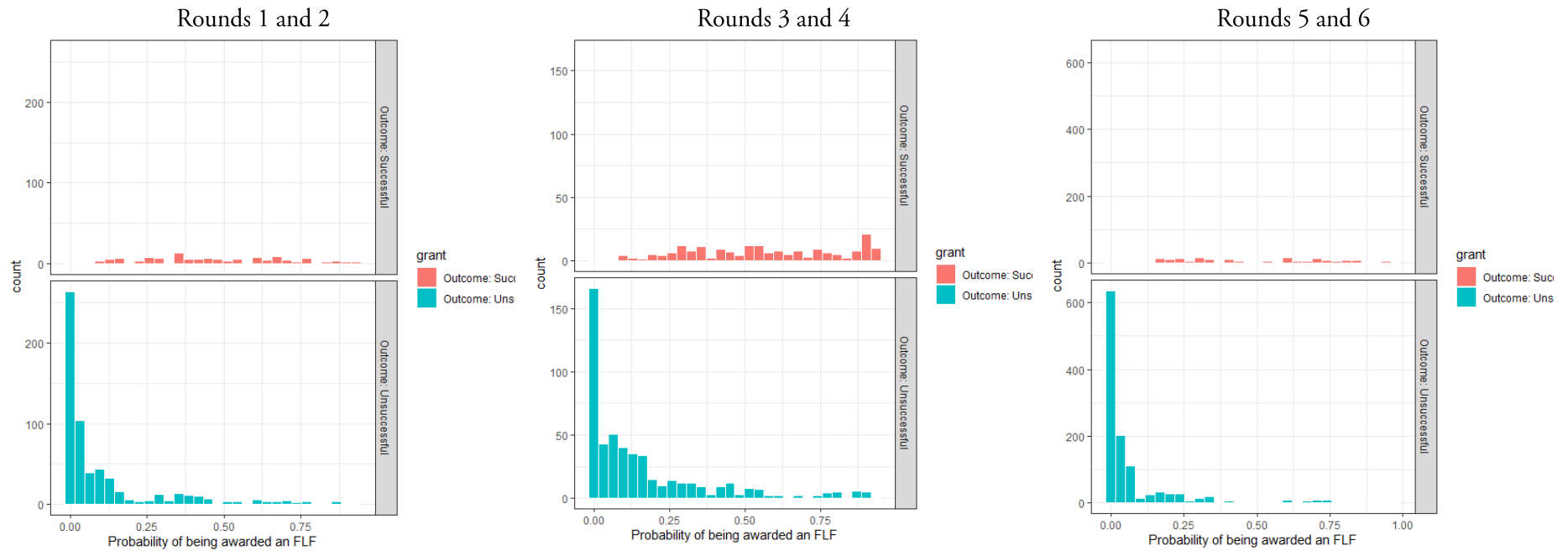


These top panel in this graph shows the distribution of propensity scores for those who were successful, and the bottom panel shows the distribution of propensity scores for those who are not awarded a fellowship. The *common support* assumption would necessitate that there is non-zero density in the top panel for all those scores that have a non-zero density in the bottom panel and vice versa. We can see here that this is no true in all rounds for those at the tail end of the unsuccessful applicants and at the top end of successful applicants. For instance, in all rounds, no successful applicant had a propensity score below 0.10, but there is a large density of unsuccessful applicants who had approximately zero likelihood of being awarded an FLF. We must thus restrict the counterfactual to not include any of these individuals as they would otherwise violate the *common support* assumption.

Source: RAND Europe analysis

The model used for PSM uses all those covariates that appear to differ significantly in any one round, but we expect that the largest single determinant of success is the SIFT score.  To generate a propensity score, we  model success using a logit model as follows:

$$\Pr(success_i) = \frac{\exp\left(\beta_1 + \beta_2 Sift_i + \beta_3 female_i + \beta_4 under40_i + \beta_5 white_i + \beta_6 UK_i\right)}{1 + \exp\left(\beta_1 + \beta_2 Sift_i + \beta_3 female_i + \beta_4 under40_i + \beta_5 white_i + \beta_6 UK_i\right)}$$

where *Sift* is an applicant's SIFT score, *female* is a dummy variable taking on value 1 if the applicant is female, *under40* is a dummy variable taking on value 1 if the applicant is under 40, *white* is a dummy variable taking on value 1 if the applicant identifies as white and *UK* is a dummy variable taking on value 1 if an individual is a UK citizen. As the coefficients of the logit model used to generate the likelihood of being successful (propensity score) are not used themselves in analysis, we do not present the results of the underlying model here.

There are a number of different techniques for matching individuals to the counterfactual group based on propensity scores. A common, and perhaps the most intuitive method, is that of nearest neighbour matching, using a 1:1 matching ration, where we iteratively select the unsuccessful applicant with the nearest propensity score without replacement. However, due to the substantially different distributions of propensity score across the successful and unsuccessful applicants, in this instance nearest neighbour matching generates a poor counterfactual, with a substantial difference in sift scores still between treatment and generated counterfactual. This is demonstrated in Table 11 below for each of the sub-samples.

Table 11: Comparing treatment with PSM-generated counterfactuals– nearest neighbour matching

|  | Sample Size | Female | Under 40 | White | UK Citizen | Sift Score |
|---|---|---|---|---|---|---|
| Rounds 1 and 2 | | | | | | |
| Treatment | 122 | 0.43 | 0.85 | 0.76 | 0.54 | 8.83 |
| Counterfactual | 122 | 0.34 | 0.84 | 0.77 | 0.48 | 8.65 |
| Rounds 3 and 4 | | | | | | |
| Treatment | 190 | 0.47 | 0.83 | 0.84 | 0.52 | 8,84 |
| Counterfactual | 190 | 0.45 | 0.85 | 0.75 | 0.52 | 7.96 |
| Rounds 5 and 6 | | | | | | |
| Treatment | 176 | 0.43 | 0.85 | 0.79 | 0.56 | 9.01 |
| Counterfactual | 176 | 0.42 | 0.83 | 0.77 | 0.54 | 8.77 |

Source: RAND Europe analysis

We instead explore different ways of generating a better counterfactual. There are number of possible ways of doing this, but most of these reduce either the sample size of the counterfactual and/or the treatment group. One method is to undertake nearest neighbour matching but with replacement, which effectively weights or matches closely matched individuals multiple times. However, this approach reduces the effective sample size of the counterfactual to as low as 33 (for Rounds 1 and 2). Most other options also effectively give greater weight to those individuals that are closely matched to the treatment group (at the higher end of region of common support). As discussed above, ultimately coarsened matching produced the closest results whilst maintaining a sufficiently large effective sample size to ensure we are not underpowered. Under coarsened matching, each covariate is coarsened into groups and then individuals are matched by covariate groups rather than exact covariate values. As an example, consider matching on date of birth: under exact matching we would look to find an individual with the same date of birth and under coarsened matching we would match individuals by month or year of birth, instead of exact birth date.

**Table 12: Metrics to assess balance of counterfactual and treatment – coarsened exact matching**

|  | Sift Score | Female | Under 40 | White | UK Citizen |
|---|---|---|---|---|---|
| **Rounds 1 and 2** | | | | | |
| Standardised difference in means | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 |
| Ratio of variances | 0.8336 | | | | |
| Mean difference in eCDF | 0.0089 | 0.000 | 0.000 | 0.000 | 0.000 |
| Maximum difference in eCDF | 0.0662 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Rounds 3 and 4** | | | | | |
| Standardised difference in means | 0.058 | 0.000 | 0.000 | 0.000 | 0.000 |
| Ratio of variances | 0.910 | | | | |
| Mean difference in eCDF | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 |
| Maximum difference in eCDF | 0.112 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Rounds 5 and 6** | | | | | |
| Standardised difference in means | 0.099 | 0.000 | 0.000 | 0.000 | 0.000 |
| Ratio of variances | 0.835 | | | | |
| Mean difference in eCDF | 0.013 | 0.000 | 0.000 | 0.000 | 0.000 |
| Maximum difference in eCDF | 0.118 | 0.000 | 0.000 | 0.000 | 0.000 |

To evaluate how well-balanced the counterfactual generated by coarsened matching, we look at various metrics to examine the distribution of counterfactual and treatment groups on all covariates. The coarsened matching approach ensures the standardised mean difference is close to zero for all covariates and the ratio of variances (calculated for continuous variables) is close to one, both of which suggest that the coarsened

matching approach does generate a sufficiently close counterfactual. The difference in the empirical cumulative distribution function is close to zero, both in terms of mean difference and maximum difference, suggesting that coarsened matching does generate an appropriate counterfactual.