

# **UKRI Future Leaders Fellowships**

## **Evaluation Scoping & Feasibility Study**

Emma Pollard, Becci Newton, Helen Gray, Clare Huxley, James Cockett,  
Dafni Papoutsaki, Institute for Employment Studies  
Professor Marc Cowling, University of Derby

# Institute for Employment Studies

IES is an independent, apolitical, international centre of research and consultancy in public employment policy and HR management. It works closely with employers in all sectors, government departments, agencies, professional bodies and associations. IES is a focus of knowledge and practical experience in employment and training policy, the operation of labour markets, and HR planning and development. IES is a not-for-profit organisation.

## Acknowledgements

The authors are indebted to Stephen Meader, Stephanie Diepeveen, Henry Cormack, Kirsty Grainger, and Jean Brown at UKRI for their steering and management; to the NPIF Evaluation Oversight Board, and the Future Leaders Fellowships Project Board for their feedback; all those who attended the workshops and provided feedback via the surveys; and to Kate Alexander, De-Jon Ebanks-Silvera, Alma Boustati and Rebecca Duffy at IES for their support with the research and report.

Institute for Employment Studies  
City Gate  
185 Dyke Road  
Brighton BN3 1TL  
UK

Telephone: +44 (0)1273 763400  
Email: [askIES@employment-studies.co.uk](mailto:askIES@employment-studies.co.uk)  
Website: [www.employment-studies.co.uk](http://www.employment-studies.co.uk)

Copyright © 2021 Institute for Employment Studies

IES project code: 5978

# Contents

<b>Executive summary .....</b>	<b>5</b>
<b>1 Introduction .....</b>	<b>10</b>
1.1 Aims and objectives .....	10
1.2 Evaluation research questions .....	11
1.3 Approach to the scoping and feasibility .....	13
1.3.1 Summarise existing knowledge .....	13
1.3.2 Consider the options .....	15
1.3.3 Synthesis .....	17
1.4 Report structure .....	17
<b>2 What does FLF aim to do? .....</b>	<b>19</b>
2.1 Introducing the Future Leaders Fellowships .....	20
2.2 Approach to developing the theory of change .....	23
2.3 Moving from the business case logic model to a theory of change .....	24
2.4 Inputs and activities .....	28
2.5 Outputs .....	30
2.6 Outcomes and intermediate impacts .....	30
2.7 Major impacts .....	31
2.8 Agents for change and the causal pathway .....	32
Mechanisms for change .....	33
2.9 From theory of change to evaluation .....	34
<b>3 What is the reality of FLF? .....</b>	<b>36</b>
3.1 How different is the FLF programme? .....	36
3.1.1 Similarity of approach .....	37
3.1.2 Potential for benchmarking .....	38
3.2 Evaluating FLF Applicants .....	39
3.3 Diversity of FLF applicants and awardees .....	44
3.3.1 Overview .....	45
3.3.2 Demographics .....	46
3.3.3 Pathways .....	51
3.4 Comparing participation in FLF .....	56
3.5 Other considerations from FLF programme data .....	57
3.5.1 Defining sub-groups .....	57
3.5.2 Consideration of near-misses .....	58
<b>4 Evaluation considerations and challenges .....</b>	<b>62</b>
4.1 Potential evaluation approaches .....	64
4.1.1 Choosing an evaluation method .....	65
4.2 Learning from other evaluations .....	67
4.2.1 Evaluation activity .....	67
4.2.2 Lessons from the evaluations .....	69
4.3 Learning from wider evaluation work .....	72
4.4 Evaluation challenges .....	73
<b>5 Outcome metrics .....</b>	<b>81</b>
5.1 Categorising the measures .....	82
5.2 About the administrative data sources .....	84
5.3 Dissemination metrics .....	86
5.4 Networking, collaboration and influence metrics .....	89
5.5 Research funding metrics .....	92
5.6 Business performance metrics – businesses only .....	93
5.7 Spin-offs/entrepreneurship metrics .....	96
5.8 Patent and IP activity metrics – STEM only .....	98

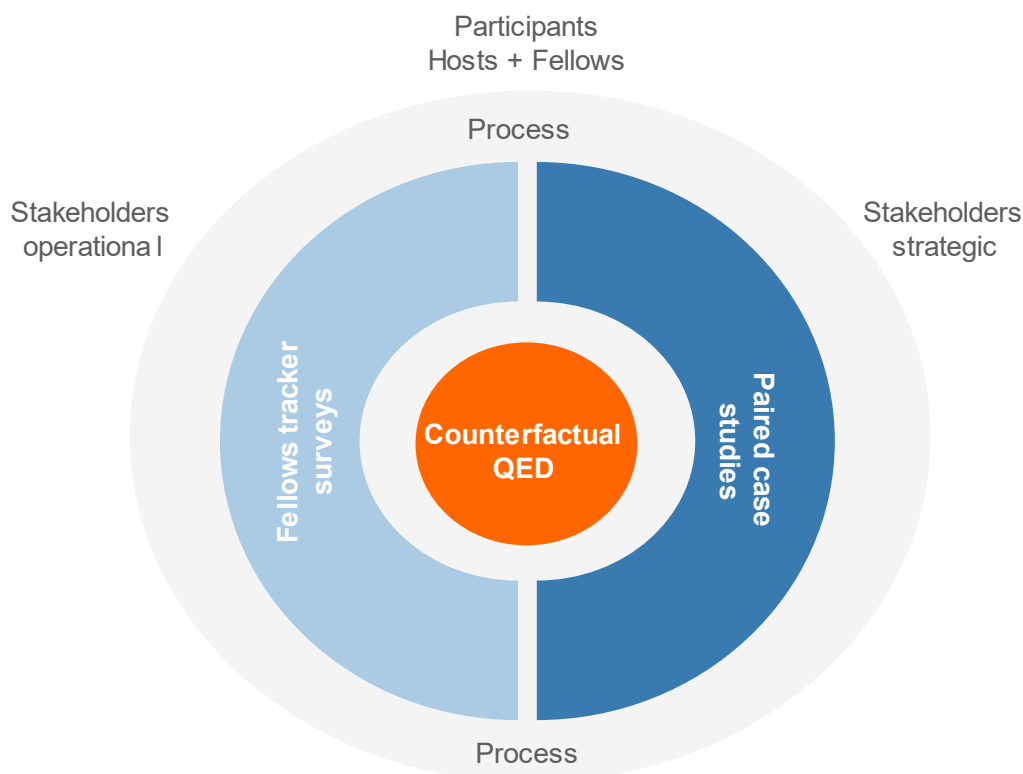
5.9	Leadership metrics .....	100
5.10	R&I metrics and wider reach .....	103
5.11	The timing of outcome metrics .....	105
<b>6</b>	<b>Considerations for a counterfactual impact evaluation .....</b>	<b>108</b>
6.1	Estimating causal impact.....	109
6.1.1	<i>The selection process and application data .....</i>	<i>109</i>
6.1.2	<i>Potential comparison groups.....</i>	<i>110</i>
6.1.3	<i>Feasibility of processing personal data for the counterfactual impact evaluation .....</i>	<i>114</i>
6.1.4	<i>Summary.....</i>	<i>115</i>
6.2	Proposed approach to estimating the impact of FLF.....	115
6.2.1	<i>Propensity Score Matching .....</i>	<i>115</i>
6.2.2	<i>Regression Discontinuity Design.....</i>	<i>117</i>
6.2.3	<i>Approach to data collection .....</i>	<i>119</i>
6.2.4	<i>Estimating impacts at different levels .....</i>	<i>120</i>
<b>7</b>	<b>Considerations for process evaluation of FLF.....</b>	<b>122</b>
7.1	What process and theory-based evaluation add .....	122
7.1.1	<i>What are process and critical realist evaluations? .....</i>	<i>123</i>
7.1.2	<i>Objectives for process and theory-based evaluation .....</i>	<i>123</i>
7.2	Analytic framework for the process evaluation.....	125
7.3	Design .....	127
7.4	Optimising the design.....	129
7.5	Tracking surveys .....	130
7.5.1	<i>Tracking fellows.....</i>	<i>130</i>
7.5.2	<i>Tracking hosts.....</i>	<i>136</i>
7.6	Matched case studies.....	138
7.6.1	<i>Feasibility of processing personal data for the process and theory-based evaluations .....</i>	<i>140</i>
7.7	Summary of recommendations .....	141
<b>8</b>	<b>Considerations for economic evaluation.....</b>	<b>143</b>
8.1	Approaches to understanding value for money.....	143
8.2	Understanding the value for money of FLF .....	144
8.2.1	<i>Establishing absolute value for money.....</i>	<i>145</i>
8.2.2	<i>Establishing relative value for money.....</i>	<i>147</i>
8.3	Summary recommendations .....	148
<b>9</b>	<b>Recommendations for a full evaluation .....</b>	<b>150</b>
9.1	What this means for the scoping and feasibility questions.....	152
9.2	The design of the FLF evaluation.....	155
	<b>References.....</b>	<b>161</b>
	<b>Annexe:.....</b>	<b>167</b>
	<b>Appendix A: Research landscape .....</b>	<b>168</b>
	<b>Appendix B: Mapping fellowship programmes.....</b>	<b>177</b>
	<b>Appendix C: Evaluation methodologies .....</b>	<b>184</b>
	<b>Appendix D: Learning from other evaluations .....</b>	<b>194</b>
	<b>Appendix E: Potential administrative data sources .....</b>	<b>213</b>

## Executive summary

The overarching aim of this scoping and feasibility study was to identify the appropriate methods to support answering the evaluation questions developed for the Future Leaders Fellowships (FLF) programme by proposing a suitable, workable approach to monitoring and evaluating the FLF programme. This extensive review has identified that a comprehensive, multi-component evaluation model - encompassing counterfactual and theory-based impact, process and strategic assessment of the programme, will be the optimal to understand the programme's effects at the individual and aggregate level.

Due to the complexity of the programme, a comprehensive evaluation is necessary because there is a need to explain not only if the programme had the planned impact, but how it had this impact. Furthermore, the robustness of the measures that are available to understand anticipated impacts vary. As a result, the programme will benefit from the combination of counterfactual impact evaluation at the individual level; theory-based impact evaluation; and process evaluation which can report at the host level as well as in aggregate for the programme.

**Figure A: Overview of evaluation model**



The key messages from this study, addressing the questions that were set by UKRI for the Scoping and Feasibility Review, are:

**Is it possible to provide a robust evaluation of the FLF, and if so when would be the most appropriate time to instigate this?**

The review has established that the degree of robustness will vary within the FLF evaluation. Nonetheless, it is possible to take forward a suite of approaches that will identify the range of impacts stemming from this complex programme.

A full, counterfactual net impact evaluation can address questions of impact at the individual, fellow level. It will provide insights into the inter-related causal pathways of (i) the research and innovation (R&I) idea and (ii) the host. Theory-based impact evaluation can detect progress along the causal pathways, alongside attitudinal and behaviour changes, to understand the effects of FLF more roundly on outputs and outcomes of Fellows, hosts and for equality, diversity, inclusion and porosity (movement within the R&I ecosystem) referred to as EDIP.

The FLF is a large, complex programme seeking ultimately to effect change on a range of measures at the wider R&I and at the societal level. These will take significant time to emerge (in all likelihood). The evaluation should continue beyond the funded period, ideally for a minimum of ten years to allow granular insights into impacts for different disciplines and multi- and inter-disciplinary research and innovation (MIDRI) mixes, as well as in different host settings (related to the recommended segmentation) to be captured. With these points in mind, a review of the intake following the award of Round 6 will mean the viability of taking forward the most robust impact measurement for all fellows or a subset related to the four segments related to host and discipline can be confirmed.

**What is the most appropriate broad, sensible and feasible approach to evaluating the FLF scheme?**

*And (i) how do measurements relate to a theory of change? (ii) how often should they be captured? (iii) and at what level should measurements be made?*

A multi-component evaluation will enable UKRI to understand (a) whether or not the programme achieved its planned impacts and indeed if any unanticipated impacts emerge, (b) for whom and in what circumstances and (c) contextual factors that supported or inhibited the emergence of impacts.

The measures scoped aim to capture outcomes identified in the theory of change – mainly at the intermediate impact level for the fellow, host, and EDIP pathways that will provide insights to the R&I pathway - while keeping in mind DORA principles and the need to minimise burden on Fellows and other sector agents and stakeholders. Administrative data sources minimise burdens and increase the chance of collecting outcomes for a large proportion of fellows and a nominated comparison group.

Primary data collection remains necessary to fully round out the picture for fellows, and to get to host outcomes, where administrative sources are more limited. Attitudinal and behavioural changes are particularly important to the fellow pathway and link strongly to the

EDIP pathway. Engaging fellows in a career tracker will capture these outcomes and help establish how their attitudes and behaviours lead (or do not lead) to the planned intermediate outcomes. Host surveys are crucial, and should aim to capture information at baseline, at the end of the initial funded period, then following the contract extension period is likely to be optimal.

Given the complexity of the programme, measures must encompass the four key segments related to host and discipline, and must be sensitive enough to pick up on MIDRI discipline mixes. The key unit of assessment for the counterfactual impact evaluation is the individual fellow. The theory-based evaluation can go beyond this to encompass hosts, EDIP and R&I and effects in aggregate.

### **Given the balance of academic/business-based fellows, how should this be reflected in any potential evaluation?**

It is crucial that outcomes for academic and business-based fellows and hosts are understood. The theory of change sets out ambitions that the FLF encourages a greater range of host organisations into R&I, that these hosts will be retained in the sector increasing the extent of positive competition and collaboration, and that in turn will lead onto a system that supports ground-breaking R&I.

The reality of FLF is that awards have been made predominantly to hosts in academia. This will limit the ability to find granular counterfactual impact for fellows in business and it presents risks in terms of business hosts not responding to primary research in theory-based evaluation. Using the matched case, theory-based method, drawing on published secondary data sources (appropriately tailored), will help to surface the impacts for these smaller segments.

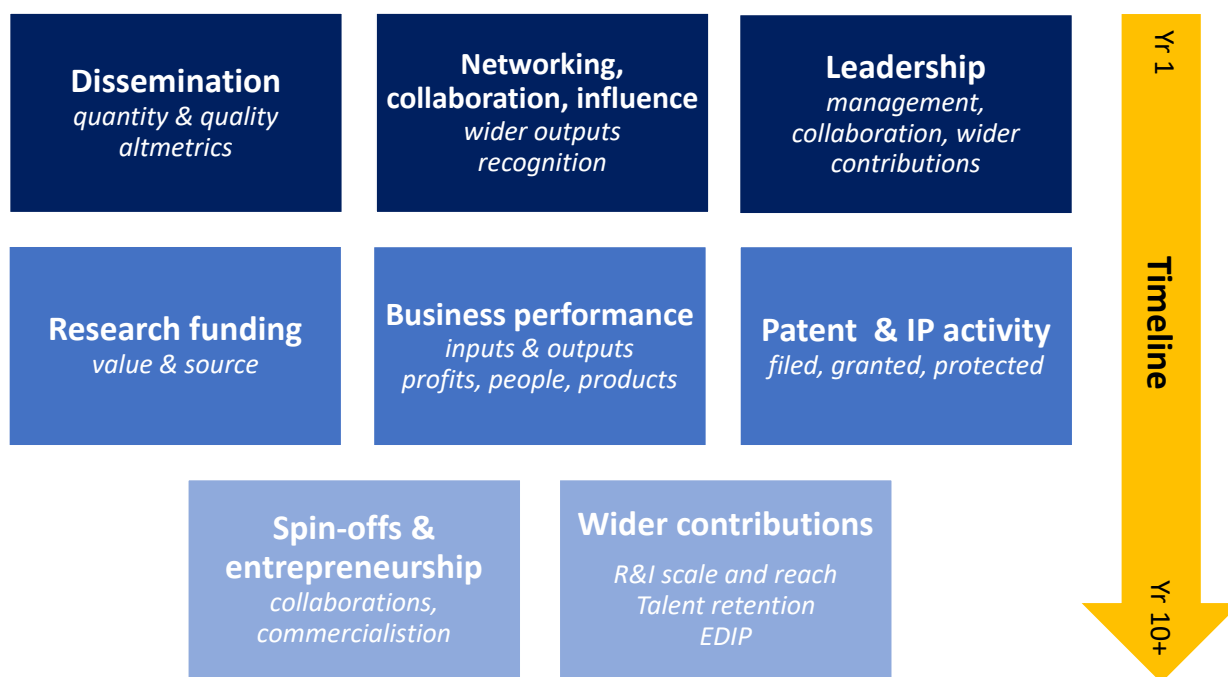
### **How, and to what extent, can the different impacts of the FLF be measured (both objectives from the business plan, and possibly also softer impacts of leadership and behavioural change)?**

This study has established the means through which hard outcomes and the softer ones can be measured. It is essential to measure both as key aspects of the theory of change could not be tested without this. The recommended approach will capture hard metrics for the fellows that are consistent with R&I but stretch beyond the narrow measures. Some of this can be used to understand PI behaviour around collaboration. Primary research with fellows and hosts will add dimensionality to this, and allow the interaction between the inputs, and outcomes to be better understood and to firm up the mechanisms through which the FLF is leading to impact.

There are limitations and constraints for the evaluation that extend beyond the methodological and instead relate to practical decisions made about evaluation. For example, it is likely that only long-term evaluation will establish the effect of FLF at the major/final impact level envisaged in the theory of change; and the extent to which granular detail on the causal pathways and interaction between them will depend on the frequency of evaluation rounds, and extent of engagement in each.

The figure below sets the overarching themes against which metrics for outcomes have been suggested. This demonstrates how the approach will cover the range of outcomes envisaged by the theory of change, but also shows the time involved for them to emerge.

**Figure B: Overview of the metric themes**



Source: IES 2021

A multi-component evaluation will ensure robust and rich information is gathered to show the impact of the programme, explain how this was achieved and to illustrate factors and facets that were important to the effects that are detected. Key approaches and how they address the evaluation questions are set out below.

**Table A: Evaluation framework**

Research issue	Method
Meeting targets for inputs and outputs	MI analysis; Stakeholder interviews
Stimulation of multi- and interdisciplinary research and innovation (MIDRI) fellowships	MI analysis; Stakeholder interviews; Case study exemplars; Peer matched case method
Additionality	Stakeholder interviews; Comparative analysis against existing evidence; Career tracker survey
Effectiveness of post award management	Programme documentation; Career tracker survey; Qualitative interviews with fellows; Stakeholder interviews (operational); Peer matched case method
What is working well/ less well	Stakeholder interviews (operational and strategic); Qualitative interviews with fellows; Qualitative interviews with hosts
Unexpected barriers or facilitators to outcomes	Stakeholder interviews (operational and strategic); Qualitative interviews with fellows; Qualitative interviews with hosts



Research issue	Method
Effective targeting of funding to meet programme aims	MI analysis; Stakeholder interviews; Career tracker survey outcomes
Lessons	Stakeholder interviews; Meta/comparative analysis
Effect on increasing high quality and impactful R&I	QED counterfactual impact; Strategic stakeholder interviews
Meeting wider government objectives	QED counterfactual impact; Strategic stakeholder interviews; Theory-based impact nested case studies
Increasing MIDRI and cross-sector working	QED counterfactual impact; Theory-based measures in fellows and host surveys; Theory-based impact matched peer case studies
Developing R&I leaders	QED counterfactual impact; Theory-based measures in fellows surveys using common metrics with other schemes; Theory-based impact matched peer case studies
Attracting and retaining international talent	QED counterfactual impact; Theory-based measures in fellow and host surveys; Strategic stakeholder interviews
Increasing the reputation of the UK as a place to pursue R&I careers	Theory-based measures in fellow and host surveys; Strategic stakeholder interviews
Increasing engagement between industry and academia for R&I	QED counterfactual impact; Theory-based measures in fellow and host surveys; Theory-based impact matched peer case studies
Developing a more equal, diverse and inclusive R&I workforce?	Theory-based measures in fellow and host surveys; Strategic and operational stakeholder interviews
Influencing risk appetite for novel R&I	Theory-based measures in fellow and host surveys; Strategic and operational stakeholder interviews; Theory-based impact nested case studies
Hosts embed planned behaviours	Theory based measures in host and fellow surveys; Theory-based impact nested case studies; Strategic and operational stakeholder interviews
Increasing R&I careers in new and novel areas	QED counterfactual impact; Theory-based measures in fellow and host surveys; Theory-based impact matched peer case studies
Contribution to new investment into R&I from outside government	QED counterfactual impact; Theory-based measures in fellow and host surveys; Strategic stakeholder interviews
Improving Host Organisations' support for early career R&Is	Theory-based measures in fellow and host surveys; Strategic stakeholder interviews
Wider, overall impact on UK R&I	QED counterfactual impact; Theory-based measures in fellow and host surveys; Strategic stakeholder interviews
Wider, overall economic impact	Strategic stakeholder interviews; Theory-based measures in fellow and host surveys
Wider, overall societal impact	Strategic stakeholder interviews; Theory-based measures in fellow and host surveys; Theory-based impact matched peer case studies
Value for money (absolute and relative)	Strategic stakeholder interviews; potentially economic assessment based on valuing outcomes derived from QED counterfactual impact and cost information

# 1 Introduction

---

This chapter sets out the purpose of the scoping and feasibility study and the methods undertaken, it also describes how the report is structured.

## 1.1 Aims and objectives

UK Research and Innovation (UKRI) was established in 2018 as the UK's national funding agency for science and research in order to join up the funding landscape and help translate excellent research into better business outcomes. It works to strengthen and promote world-leading research and innovation (R&I) across and between all disciplines. One of its strategic priorities is to lay the foundations for excellent R&I by nurturing the pipeline of current and future talent. Under this Leading Talent agenda, UKRI created their new Future Leaders Fellowships (FLF) programme drawing on the National Productivity Investment Fund (NPIF) to develop the next generation of innovation and research leaders.

UKRI requires a rigorous evaluation of its Future Leaders Fellowship programme to help understand its longer-term impact. However there are a number of challenges in attempting to identify and quantify the impact and value of such programmes: attribution of impact, diversity of population/breadth of scope (segmentation), determining success and defining terms, varying levels of potential impact (individual, teams, organisation, society), identifying additionality (counterfactual), capturing wider benefits and unintended outcomes (spillovers), taking account of external influences, harnessing existing data, timing of data capture, and a need for intermediate outcome indicators.

UKRI commissioned IES to conduct this Scoping and Feasibility Study to help to identify appropriate methods to support answering the FLF evaluation questions by proposing a suitable, workable approach to monitoring and evaluating the FLF programme. This study had three key objectives:

- i) to develop a summary of existing knowledge and information about potential evaluation methods,
- ii) to explore the challenges around evaluating the Future Leaders Fellowship programme (including the trade-offs between evaluation approaches),
- iii) to make recommendations for the full evaluation.

The key questions for the scoping and feasibility study were:

1. Is it possible to provide a robust evaluation of the FLF, and if so when would be the most appropriate time to instigate this?

2. What is the most appropriate broad, sensible and feasible approach to evaluating the FLF scheme
  - i. How do measurements relate to a theory of change?
  - ii. How often should they be captured?
  - iii. At what level should measurements be made?
3. Given the balance of academic/business-based fellows, how should this be reflected in any potential evaluation?
4. How, and to what extent, can the different impacts of the FLF be measured (both objectives from the business plan, and possibly also softer impacts of leadership and behavioural change)?

## 1.2 Evaluation research questions

UKRI set out questions for the evaluation to address, developed through an internal review process, which have guided this scoping and feasibility study and its recommendations (see Table 1.1). The evaluation questions seek to establish if the FLF has an impact, if so, how and through what means, and the value for money this represents. This leads towards a design that can address counterfactual and perceived impact, and capture and document programme delivery to understand why these impacts emerge.

The way in which questions are framed indicates UKRI's dual objectives to not only surface whether or not the programme achieves its planned impacts, but if it does, how it does this – therefore a strong case is made for impact analyses alongside theory based and process evaluation to provide the necessary explanation. Questions covering 'to what extent' and 'how' are best addressed through narrative research whereas questions exploring 'if' and 'whether' can be tackled using counterfactual approaches.

**Table 1-1: Evaluation questions**

Process evaluation	Sub-questions
<b>Appropriateness</b> To what extent (and how) is the FLF programme working and being delivered as intended?	<ul style="list-style-type: none"> <li>- (To what extent) are targets for FLF inputs and outputs being met?</li> <li>- (To what extent) has FLF stimulated MIDRI fellowships?</li> <li>- (To what extent) are FLF fellowships additional to other schemes supported by UKRI and other UK / international funders?</li> <li>- (To what extent) has FLF delivered effective post award management to support the professional development of the fellows?</li> <li>- What, in practice, is felt to be working more / less well regarding the delivery of the fund, and why?</li> <li>- What are the unexpected barriers or facilitators to the FLF processes and the delivery of the anticipated outcomes, if any?</li> <li>- 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?</li> <li>- What lessons are there for future rounds / similar schemes?</li> </ul>

Impact evaluation	Sub-questions
<b>Efficiency</b> (How) has FLF changed the research and innovation landscape?  To what extent, and how, has FLF delivered highly skilled research and innovation leaders of the future?  To what extent, and how, does FLF make the UK attractive place for future R&I leaders?  (How) has FLF led to a change in behaviour for early career researchers, innovators and hosts?          To what extent (and how) has the FLF delivered wider knowledge, economic and societal impacts?	- (To what extent) has FLF increased high quality and impactful research and innovation? - (To what extent) and how does the overall FLF fund support wider government objectives? - (To what extent) has FLF increased MIDRI and crosssector working, for the Fellow, Fellow's team and / or collaborators?  -(To what extent) has FLF developed R&I leaders over and beyond other fellowship (or similar) schemes?  -(To what extent) has FLF developed attracted and retained talent (fellows and associated teams) to the UK? - To what extent, and how, has FLF influenced the reputation of the UK as a place to pursue a career in research or innovation?  -(To what extent) has FLF increased engagement between industry and academia on research and innovation activities? -(To what extent) has FLF developed a more equal, diverse and inclusive research and innovation workforce? -To what extent, and how, has FLF influenced the appetite to risk to novel R&I of early career researchers and innovators, panel members and hosts? -(To what extent) have host organisations promoted and supported the FLF scheme and delivered against expectations / commitments for research or innovation support? -(To what extent) has FLF increased careers in research and innovation within new and novel areas? -(To what extent) has FLF contributed to new investment into research and innovation from outside government? -(To what extent, if any) has FLF influenced, or set precedents for, improvements in Host Organisations' support for early career researchers or innovators, EDI or related UKRI policy goals?  -What has been the wider, overall impact of the FLF on UK research and innovation expertise and on other parts of UKRI practice? -What has been the wider, overall economic impact of the FLF, including the economic value of non-market impacts? -What has been the wider, overall societal impact of the FLF, including the impact on the number of high-quality jobs, wages, and wellbeing, and societal benefits from new products and services?
Economic evaluation	Sub-questions
<b>Value for money</b> Based on the overall, estimated impact of the FLF – considering those impacts which can be	-To what extent does the FLF represent value for money in absolute terms? -To what extent does the FLF represent value for

given market and non-market values – compared to the overall cost of delivering the FLF, to what extent does the FLF represent value for money?

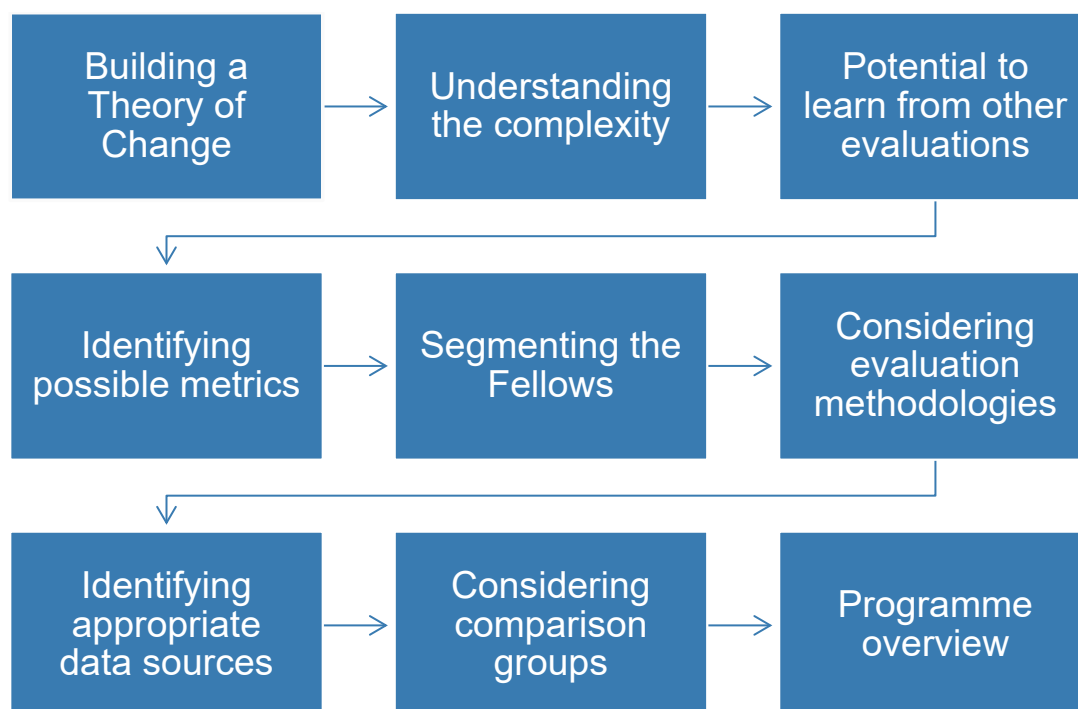
money compared to other possible alternative ways of achieving the same impacts?

Source: UKRI, 2020

## 1.3 Approach to the scoping and feasibility

This section outlines the approach and methodology that underpins the findings presented in this report.

**Figure 1-1: Visualisation of scoping and feasibility approach**



### 1.3.1 Summarise existing knowledge

The first stage had two aims: to obtain an overview of evidence from existing and prior evaluations and fellowship programmes, and to develop a Theory of Change (TOC) for the FLF to lay out its distinctive approach and anticipated impacts. This involved a review of the literature on evaluation approaches and on fellowship and innovator programmes (and any evaluations of these) and work with key stakeholders to develop a TOC.

Monitoring and evaluation approaches were reviewed to identify potential methods and assess their appropriateness in terms of: the settings/context of use, benefits afforded, potential limitations and the conditions required to be most effective. The aim was to draw on best and emerging practice from research and policy evaluation including HM Treasury's Green Book (2018) and the Magenta Book (2020), and Innovate UK's

Evaluation Framework. It also aimed to identify methods and approaches used in different settings, in academia, business and policy; and in different disciplines; new methods and approaches as well as established practices.

Additional desk research was undertaken to identify programmes, schemes and initiatives aimed at developing research capability in the UK, supporting early career researchers (ECRs), and developing research leaders. This aimed to build understanding of the volume, scope, nature and goals of these programmes, given potential overlap in anticipated impacts with the goals and ambitions for the FLF programme. This is critical to understanding the feasibility of attributing impact to FLF (a key challenge). To identify and map relevant programmes the team looked for schemes with one or more of the following attributes:

- Programme offering support/funding/pay for at least one year;
- Possibly offering a number of awards/fellows (rather than a one-off award/prize);
- Located in academia, business or wider including third sector organisations and government;
- Aimed at those with a PhD and/or in their early stages of a research career;
- Aimed at supporting career development and/or leadership;
- Hosted mainly in the UK.

To search for potential programmes (including those closed, currently in operation and in development), the team reviewed the websites of known key funders including UKRI and the individual research councils, key HE sector bodies such as UUK, and websites dedicated to supporting researchers. The latter included: Vitae (the key organisation supporting the professional development of researchers), EURAXESS (a pan-European initiative delivering information and support services to professional researchers), Early Careers Researchers Central (a platform for the early career researcher community), and Research Professional.

Identified programmes were reviewed and details of their name, funder, aims, eligibility, support (particularly funding provided), duration, volume were recorded. The team also identified where there was evidence that programmes were monitored and/or evaluated and recorded the approaches used, their contexts, how practice has changed over time, and areas of good practice.

The scoping and feasibility work also sought to develop a TOC for the FLF programme. This builds on a Logic Model for the programme developed by UKRI alongside its original business case. A Logic Model establishes programme inputs, activities (including intended beneficiaries), outputs, outcomes and desired impacts; aims to capture how a programme should lead towards its intended impacts from an operational perspective. A Theory of Change captures the picture at a strategic level and seeks to demonstrate how and why an intervention should lead to the planned impacts. It considers the contextual factors that will have an interplay with impacts, the assumptions underlying the programme logic and the distinct causal pathways from different inputs, outputs and outcomes to final impacts.

Logic Models and Theories of Change help to refine and prioritise research questions, and identify 'causal pathways' (the set of processes through which an outcome is brought into being), which in turn help to identify what to measure to understand short, mid and longer term effects and sources of evidence. The scoping and feasibility work produced a detailed TOC that updated the Logic Model to take account of improved understanding and lessons learnt from operation of the programme (two years on). The research team convened three virtual iterative workshops with UKRI stakeholders involved in programme design and implementation and reviewed programme documentation. These activities sought to:

- Explore assumptions underpinning the programme, interdependencies and causal linkages between different aspects of the logic model, bridges (behaviours/conditions) for achieving outputs and outcomes and factors threatening their achievement.
- Look to agree what is meant by outcomes such as 'greater effectiveness', 'greater agility', and other ambitions and explore how these can best be evidenced.
- Determine the different levels of impact and nature of impact – such as individual Fellows (behaviours, knowledge base, research output, and careers); collaborators and colleagues; host organisations; research sector and workforce (and related policy); and economy and society.
- Identify steps between input and outcomes/impacts to tease out the potential for intermediate measures for establishing the potential impacts which are likely to take a long time to emerge (long term goals such as increases in research and innovation-led growth, increases in high quality jobs/wages, societal benefits from new products and services, and increased research and innovation capability for the UK).
- Develop and refine a visual interpretation for the Theory of Change and an accompanying narrative.

### 1.3.2 Consider the options

The second phase of this study reviewed the challenges facing a full evaluation of the programme and considered potential approaches to mitigate or try to reduce their influence on any evaluation design. This built on the work already undertaken by UKRI, lessons from other evaluations and wider literature; and sought to take account of changing context and understanding gained from delivering the programme. Feedback from stakeholders was also sought to gain their views and perspectives of the challenges facing an evaluation of a large-scale complex programme operating over various sites and geographies, involving diverse participants, with ambitious goals and multiple potential levels of potential influence and impact. To gather these insights a consultation-based survey was developed and emailed by the UKRI FLF team and by IES to 55 to 60 stakeholders. These included: UKRI members (FLF programme team, data and evaluation team, FLF panellists, and policy staff) and external stakeholders from research councils and other funders, and other impact evaluation specialists and analysts, researchers and innovators. The survey had 8 largely open text questions, it ran for three weeks from late January 2021, and gained 23 responses.



During this stage the team also analysed management information gathered on the FLF programme participants and applicants to understand the overall programme population and characteristics of each cohort and to provide baseline data for participants. These data included: background characteristics including nationality, gender, age, ethnicity and disability; qualifications particularly PhD and time since completion of PhD; FLF application process and decision including panels and relevant scores; current location characteristics including geographical location and category of host organisation (academia, research institute, business, third sector, government etc) and current field of research (indicated by application sift panel and interview panel). The analysis also sought to understand the similarities and differences between unsuccessful applicants and those awarded a fellowship to help to determine the potential to draw a counterfactual group or number of groups from the unsuccessful applicant pool. Understanding the characteristics of successful applicants is also useful in assessing whether a wider pool of individuals in their early research careers could be used to develop a counterfactual or comparison group.

The scoping and feasibility study also involved work to identify areas for potential measures for outcomes and impacts of the programme and the data sources that could be used to provide evidence for the measures.

The measures would need to take account of the potential level of impact: individual fellows, their host organisation (academic institution or business) and the wider teams the fellow works with, and other relevant communities and potential beneficiaries; and timeframe, to capture intermediate outcomes and impacts as well as those that may be achieved in the medium to long-term. Additionally, as the FLF programme has two fundamentally different entry pathways and sites of influence (academic and business), it was important to define impact and outcome metrics that would reflect the unique environments fellows in each pathway face; and to consider subjective or soft outcome metrics such as behavioural and attitudinal change, as well as objective hard outcomes. It is widely recognised that research is typically measured through academic metrics such as publications and citations whereas applied research is often measured by licensing income, patents and other measures of technology transfer. These are easy to measure but can be influenced by levels of scientific funding and infrastructure, may only capture a small portion of innovation activity and can ignore wider economic and social outcomes (UKRI SAID, 2018). It was important for the scoping and feasibility study to consider a broad range of potential measures to cover both academic and applied research, and also to cover researcher development.

A range of data sources were scoped for the potential to produce suitable measures to try to understand issues around access, timeliness, scope and coverage, limitations and caveats. Potential data sources included:

- Publicly available national and international data which included administrative and survey-based data on graduate outcomes, research and innovation outcomes, academic and industry interaction, and business outcomes.
- Data sources that emerge through delivery of the programme including management information captured during application and on programme (eg outcomes records, annual meetings, performance and monitoring, review reports).



- Primary data captured through fellow surveys and interviews (along the lines of the career tracker surveys adopted by the Wellcome Trust and The Royal Society)
- Wider contextual data to aid understanding of externalities and that could be factored into impact calculations such as disciplinary differences in research outcomes, and the wider context at entry for each cohort to understand the baseline condition; and contextual changes which could also explain changes in outcomes and impacts.

The work on metrics and data sources drew heavily on the Theory of Change, analysis and understanding of programme data (and data collection processes), lessons from other evaluations and evaluation frameworks and stakeholder consultation. Stakeholder consultation involved facilitating a second set of workshops with key stakeholders. These sought to discuss what success would look like for individual fellows and the programme itself, to assess potential metrics and the practicalities of using these in an evaluation including appropriate timeframes for measurement, and to identify possible data sources. A second consultation survey was also undertaken which aimed to critique a suggested set of secondary data sources and canvass suggestions for additional potential data sources to measure outcomes and impacts of the programme. The survey had 8 open text questions, ran from mid-March 2021 for two weeks and gained 9 responses.

### 1.3.3 Synthesis

The final stage of the scoping and feasibility study involved developing a shortlist of evaluation options and outlining the pros and cons of each in relation to the FLF programme in terms of ease of application (resources required, accessibility of data etc), robustness of evidence, degree to which it addresses the evaluation questions. It also involved identifying the limitations and gaps in evidence remaining, and any additional associated challenges. This stage drew on feedback from UKRI FLF leads, the NPIF Evaluation Oversight Board and the FLF Evaluation Advisory Group.

## 1.4 Report structure

The remainder of this report sets out the findings from the review in relation to the FLF programme and how these have been used to assess the options for the full evaluation and make recommendations for the approach to take.

- Chapter 2 sets out the ambitions for the FLF programme and presents the Theory of Change for FLF setting out the anticipated outcomes, intermediate impacts and wider impacts, and the mechanisms for change. This provides the basis for the design of an evaluation of the programme.
- Chapter 3 considers the reality of FLF - how FLF compares with wider fellowships and programmes aimed at supporting early career researchers, the FLF approach to assessing applications, the diversity of FLF applicants and those awarded fellowships, and explores the implications for designing an evaluation.
- Chapter 4 considers options for evaluation, methods used in evaluations of other fellowship or similar programmes and what can be learned from these studies for an

evaluation of FLF, and the challenges in evaluating the FLF programme and how they could be addressed in designing an evaluation.

- Chapter 5 considers the measures that could be used in an evaluation of FLF, following the areas set out in the Theory of Change and taking account of the reality of FLF, and identifies the most appropriate data sources, level of analysis, and timing of each and which aspects of an evaluation the measures are most suited to.
- Chapters 6, 7 and 8 set out and discuss the options and recommendations for an evaluation design:
  - Chapter 6 focuses on the design for an quasi-experimental counterfactual impact evaluation,
  - Chapter 7 focuses on the design for theory-based evaluations to include a process evaluation, a tracker survey and a matched case studies, and
  - Chapter 8 focuses on considerations for an economic evaluation for FLF.
- Chapter 9 concludes the report with recommendations for an evaluation model for FLF.

The report annexe contains a series of appendices providing further detail of the supporting evidence.

## 2 What does FLF aim to do?

This chapter sets out the ambitions for the Future Leaders Fellowships (FLF) programme, and the key elements and assumptions that the theory of change (TOC) seeks to encapsulate. The theory of change provides a guide to the considerations of metrics to be gathered to demonstrate programme effects and impacts.

### Key findings for an evaluation design for FLF:

- FLF is an ambitious and complex programme backed with significant and sustained funding. It aims to bring about a wide range of impacts through developing diverse R&I leaders and supporting diverse R&I ideas unbounded by discipline or sector. It has multiple pathways and mechanisms to achieving planned impacts. The importance of the fellow is foregrounded, and fits with the UKRI and government's R&I talent strategy, as developing influential and talented R&I leaders will enable improvements to the research ecosystem to continue beyond the initial funding period and beyond the funded R&I idea.
- The programme is still developing so evaluation can be built in from an early stage rather than designed and implemented retrospectively. However, the key architects of the programme, who also developed the original business case logic model, have moved on so the understanding and assumptions underpinning the programme are likely to have shifted reflecting the realities of delivering the programme.
- In constructing an agreed theory of change for the FLF programme, there remains some uncertainty around how the programme elements will lead to impacts (causal pathways) and thus some mechanisms to impact require further articulation. The evaluation will need to provide evidence on this and investigate the specificities of some mechanisms rather than simply tracking causal pathways.
- Key aspects of the theory of change concern there being four causal pathways to the intermediate impacts, which then lead onto the final impacts on society, the economy and related to the industrial strategy.
- The four causal pathways are (i) fellows themselves, (ii) their hosts, (iii) their R&I ideas, and (iv) equality, diversity, inclusion and porosity (EDIP). The term EDIP seeks to serve as a marker for the focus on the non-traditional in the FLF – from encouraging fellows from a more diverse range of backgrounds who are interested in multi- and interdisciplinary research, to hosts in non-traditional settings and from a greater range of academic and business settings.
- Each of the pathways leads to outcomes that then interweave in ways that cannot yet be defined to create final impacts. There are assumptions that fellows and hosts will want to build on the programme ethos of EDIP and amplify messages around the benefits of this approach, as well as embed behavioural change that produce more supportive circumstances for inclusive and porous approaches for R&I. The assumption is that breaking traditions will lead to increased innovation and better outcomes.

## 2.1 Introducing the Future Leaders Fellowships

The FLF programme was announced in 2018 and is a key element of UKRI's Talent Strategy. It is a large, unified programme aimed at helping to establish the careers of world-class research and innovation leaders across UK business and academia in order to improve innovation activity in the UK, foster and retain talent in the UK, and develop new career pathways. It aligns with the aims of the Industrial Strategy (2017) to boost productivity with investment in skills, industry and infrastructure, and through helping to translate excellent research into better business outcomes; and recipients work in the Grand Challenge areas outlined in the Industrial Strategy.

*'The aim of the scheme is to develop a research or innovation capability and person, that is nationally or internationally leading edge, and goes beyond what other, smaller or project grants could support. The fellowship may be made up of a single programme of work, or through multiple consecutive/ concurrent interlinked projects led by the fellow (a 'Portfolio Fellowship'). Where multiple shorter projects are spread over the 4 or 4(+3) years, applicants should ensure the proposed research and/ or innovation is of sufficient scale and ambition to maximise the potential and career prospects of the fellow' (Round 6 Guidance for Applicants, UKRI<sup>1</sup>, p6)*

The FLF programme is one of numerous fellowship or scholarship programmes funded and supported by UKRI (and its constituent research councils/bodies) and is part of the wider set of programmes to support and nurture talent. These include publicly funded schemes with related objectives such as the Small Business Leadership Programme, Knowledge Transfer Partnerships (KTPs), SMART awards, UKRI New Investor Grants, UKRI Impact Acceleration Accounts; and programmes supported by other bodies such as the Royal Academies, and British Academy (see below).

The programme is distinct among these many fellowship schemes as it:

- **Is not limited to specific disciplines** – the programme is intended to support individuals in all areas of UKRI's remit including multidisciplinary and interdisciplinary research and innovation (MIDRI) projects. Individuals can apply from any discipline and the research and innovation funded can cross disciplines, and the programme actively encourages cross/interdisciplinary research.
- **Has a wide(r) diversity of host organisations** and supports early career researchers working outside of academia as well as those within academe. Fellows can be located in commercial businesses as well as academic organisations (including Higher Education institutions, independent research organisation (IRO), Catapults, and Institutes). The funded work of fellows could cover innovation, incubator activity as well as research. To be successful, fellows must demonstrate that they have significant support from their hosting organisation (this commitment is part of the assessment process) which will include dedicated time on their FLF activities and thus reduced

<sup>1</sup> <https://www.ukri.org/wp-content/uploads/2021/02/UKRI-030221-FLF-Guidance-for-Applicants-Round-6-Final.pdf>

requirements for teaching or other business activities and a commitment to offering an open-ended position at the end of the fellowship. All hosts must be UK-based.

- Is **large in size** - in terms of both funding amount and numbers of awards. The total funding for the programme is £900 million and the target number of fellows is 550. Each individual fellow is generally able to apply for up to £1.5m (although there is no limit to the amount that can be requested) in a personal award. The programme funds 100% of the recipients time so covers personal and professional development as well as research project activity. Across Rounds 1 to 4, 298 Fellowships have already been awarded. At the time of writing this report, Round 5 was currently being assessed, and Round 6 (final round) was open for applications.
- Offers a **long period of funding** – the funding and wider support is available for up to seven years. The funding is provided over four years initially with the possibility of extending funding and support for a further three years after a further assessment exercise.
- Has a **wide(r) eligibility criteria** – eligibility for funding is wider than found in many early career researcher programmes, as applicants are not required to hold a PhD (to reflect the wider host organisations the programme aims to involve) nor have achieved a PhD within a set period or to hold a permanent or open-ended position. Applicants can be UK-nationals or come from abroad (with international applicants eligible for a Global Talent visa under the exceptional promise category). Additionally, the programme itself is designed to be flexible in order to attract a wider pool of applicants including those wishing to work part-time.
- Has **flexibilities** which acknowledge the realities of researchers lives and business priorities, this includes the potential to change the direction of the fellowship or alter aspects of work packages after the award has been made (eg in response to new discoveries or techniques, build on changing experience, or to reflect business or market opportunities) and to change host organisation.

The FLF programme therefore recognises the diversity of research locations and importance of interdisciplinary work (as advocated in the Nurse Review to allow ready transfer of ideas, skills and people and improve effectiveness of the research system). It recognises the diversity of research goals - to generate knowledge and innovation ('upstream'), support public policy and the needs of society, and develop useful commercial applications ('downstream'). It also provides up to seven years funding and thus recognises that 'delivering the highest quality research is difficult, requiring patience, persistence and long-term investment' (Nurse Review guiding principles, 2014, p1). It builds capacity upstream by developing talented individuals and engaging them in the research process and allows for downstream benefits to emerge – either directly through the FLF business pathway or indirectly through the academic pathway where new knowledge and innovation solutions are created, then diffused and commercialised throughout the wider ecosystem.

FLF is aimed at early-career researchers with eligibility extended to researchers with a track record of producing original research or innovative outputs rather than being focused on prior qualification criteria (such as holding a PhD). The programme is intended for researchers and innovators who have not yet secured their own funding or not yet managing their own programme of work and have the potential to be an outstanding leader

in their field. The scheme provides individual fellows with time, training, mentoring, professional development and networking; and allows them to conduct a specific programme of research or innovation (either a single programme of work or a series of interlinked projects) to work on difficult and novel challenges at the boundaries of traditional disciplines. It also provides UK host organisations including universities, research institutes and businesses with funding for up to four years, with the potential to extend for a further 3 years (up to a maximum of seven years in total), to attract, develop and retain research talent. Aspects of the FLF programme, such as networking between and across cohorts, facilitates not only the development of the breadth of knowledge and innovation, but also speed of knowledge diffusion and the spread across the ecosystem.

### Box 1: FLF programme aims and objectives.

The **aims** of the FLF programme are to produce:

- High quality and impactful R&I in areas aligned with the Government's Industrial Strategy areas
- Increased engagement between industry and academia on R&I activities including through collaboration and problem-solving
- Increased multi- and interdisciplinary R&I (MIDRI)
- Highly skilled R&I leaders of the future, from the UK and overseas
- A more equal, diverse and inclusive R&I workforce, which welcomes international talent
- Value for money.
- The **objectives** of the FLF programme are to:
  - Develop, retain, attract and sustain R&I talent in the UK
  - Foster new R&I career paths including those at the academic/business and interdisciplinary boundaries, and facilitate movement of people between sectors
  - Provide sustained funding and resources for the best early career researchers and innovators
  - Provide long-term, flexible funding to tackle difficult and novel challenges, and support adventurous, ambitious programmes.

### Importance of talent within UK policy agenda

The FLF programme sits within a wider UK research landscape that values its core ambitions of supporting and nurturing early career researchers, of promoting collaboration and multi-disciplinary and cross-sector working, and of developing leaders to support UK R&I goals but also improve the research culture. It also sits within a wider context where evaluation is expected but recognised as challenging, so the plan to build in evaluation into the FLF programme will be welcomed. Policy makers, sector bodies and key stakeholders:

- Recognise the importance of R&I to the UK's economic performance, to delivering societal benefit and to tackling domestic and global challenges including the global COVID-19 pandemic; set stretching ambitions for R&I to try to maintain the UK's strong global position against faster growing countries; and the Government has committed



significant public and private funding over the coming years alongside a plan to revitalise the R&I system in the UK.

- Demonstrate a growing interest in R&I talent and the role of highly educated and talented graduates and postgraduate employees in the innovation process is well documented. There is a shared understanding of the need to attract, train and retain diverse talent from all backgrounds including promoting the UK as a destination for talent; whilst recognising that careers in research can be lower paid and unstable; career pathways can be unclear; there is little support for movement between academia, industry and the public and third sectors or support for movement between countries; and the research culture in the UK needs attention particularly in terms of equality, diversity and inclusion but also collaboration, knowledge-sharing and collegiality.
- Understand the need for inspiring leaders who can nurture and develop future talent, which is leading the government to rethink the funding and wider support for postgraduate research and early career researchers to enable them to progress in careers inside or outside of academia
- Want to spend their money wisely and demonstrate clearly the impact (and effectiveness) of funding and research at various levels, including economic and societal; whilst at the same time recognising the importance of freeing the system from burdensome bureaucracy (See Annexe, Appendix A)

## 2.2 Approach to developing the theory of change

Theory of Change (TOC) is used to map connections between activities and outcomes, in order to generate hypotheses about how an intervention will achieve the change its designers believe will emerge. It makes explicit underlying assumptions about how programmes are expected to work (Rogers et al., 2000) to support a systematic focus on explaining how and why an intervention works (or does not). Interventions are informed by assumptions about the ways in which they will work to deliver the planned 'change'. However, in practice some of these assumptions may not hold true and/or the intervention may work in a different way than was originally intended.

Theory of change provides the means to assess these strategic and practical implications for interventions. This approach typically commences by eliciting information and developing a model for how an intervention is meant to work at intervention design stage, and then tests whether it does work in the planned ways in practice. It thus identifies gaps between the planned and actual implementation and effect(s), and explores the extent, nature, causes and resulting impacts on outcomes.

To elicit the theory of change for FLF, we led two workshops and a follow-up meeting with stakeholders within UKRI which, alongside desk research, generated the main assumptions operating within the FLF programme. It should be noted that the programme was designed at the stage the business case was submitted for funding; and that by the time of the theory of change workshops, key architects involved in the design no longer worked at UKRI and could not be involved in the process. Instead, the workshops engaged

stakeholders involved in various aspects of delivery of the programme and its planned evaluation.

The first of the workshops covered stakeholders' expectations for delivery and outcomes and how these would lead to impact. Following the first workshop, a draft 'straw man' theory of change was devised. This was discussed and tested at the consensus-building second workshop where the key assumptions were explored further and clarified along with the interdependencies and causal linkages. The draft was then reworked and confirmed in the smaller stakeholder meeting.

The consultations were followed by a multi-stage drafting process with the lead analyst at UKRI to arrive at this description of the TOC and how the programme should work.

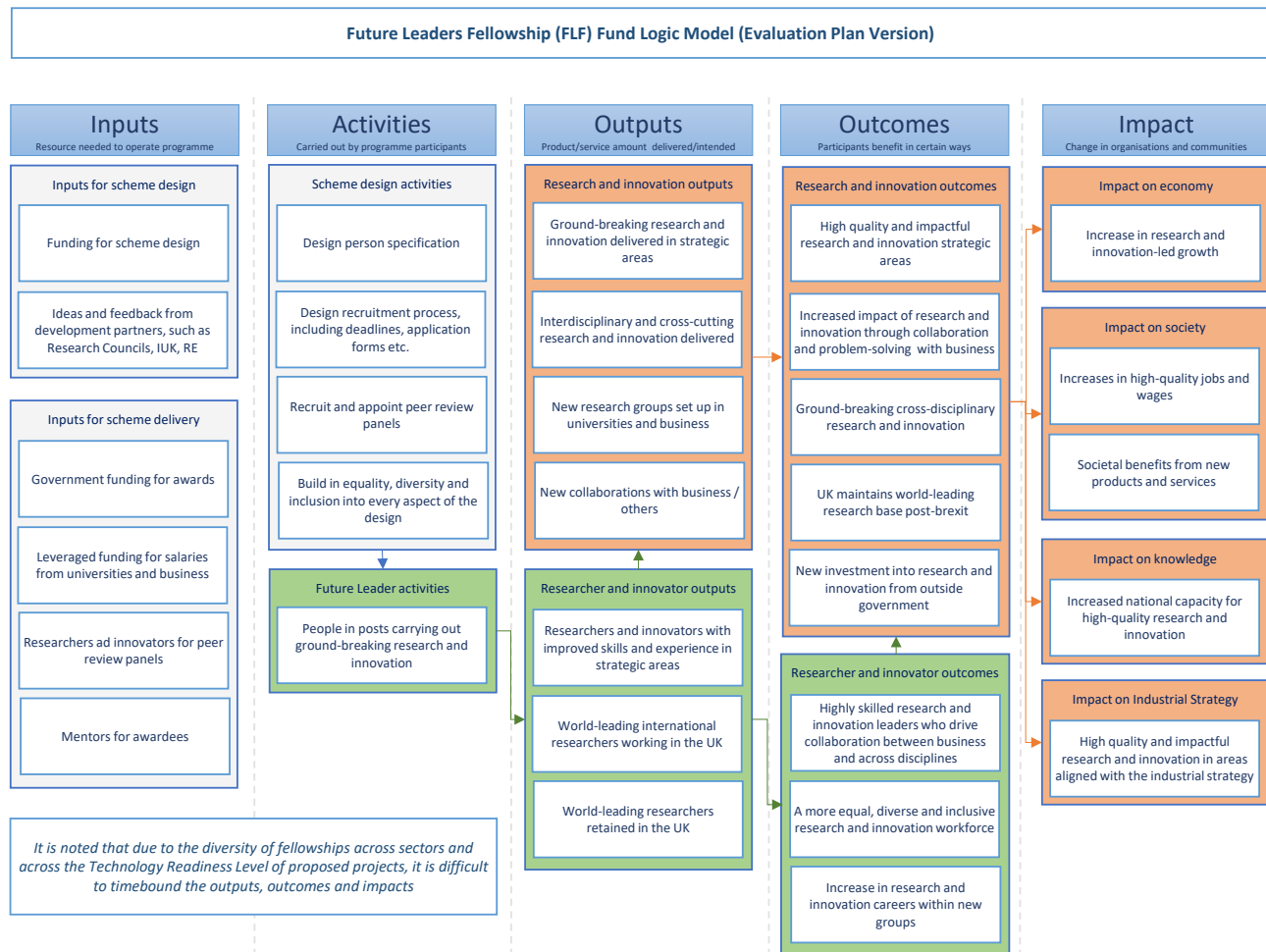
As noted, the aim of the workshops was to build consensus on the model and the assumptions about how the FLF programme worked within UKRI; while this was achieved, there was not in all cases strong certainty for how programme elements would lead to impacts and the future evaluation will need to provide evidence on this. The FLF TOC is intended to act as a benchmark for the programme, its implementation and impacts, and aspects of the causal pathways it records will be tested within the evaluation.

The following description outlines the intended TOC for the FLF programme, acknowledging that the assumed mechanisms and pathways will need to be tested as part of the evaluation.

## **2.3 Moving from the business case logic model to a theory of change**

The starting point for the development of the TOC was the logic model UKRI created when it submitted a business case for funding (see Figure 2-1).



**Figure 2-1: The FLF business case logic model**

Source: UKRI, 2020

This indicates that the fellows are intended to undertake ground-breaking research across disciplinary areas, and work across academia and business, building the capacity and capability of the UK's world-leading research and innovation base. It also identifies the major impacts from the programme that will emerge in respect of the economy, on society and knowledge and that align to the industrial strategy. Outcomes that would lead to these were primarily driven by the R&I with fellows' outcomes a precursor to these (see Box 2).

### Box 2: The R&I outcomes are:

- High quality and impactful R&I strategic areas
- Increased impact of R&I through collaboration and problem-solving with business
- Ground-breaking cross-disciplinary R&I
- UK maintains world-leading research base post-Brexit
- New investment into R&I from outside government.

The precursor fellow outcomes are:

- Highly skilled research and innovation leaders who drive collaboration between business and across disciplines

- A more equal, diverse and inclusive research and innovation workforce
- Increase in research and innovation careers within new groups.

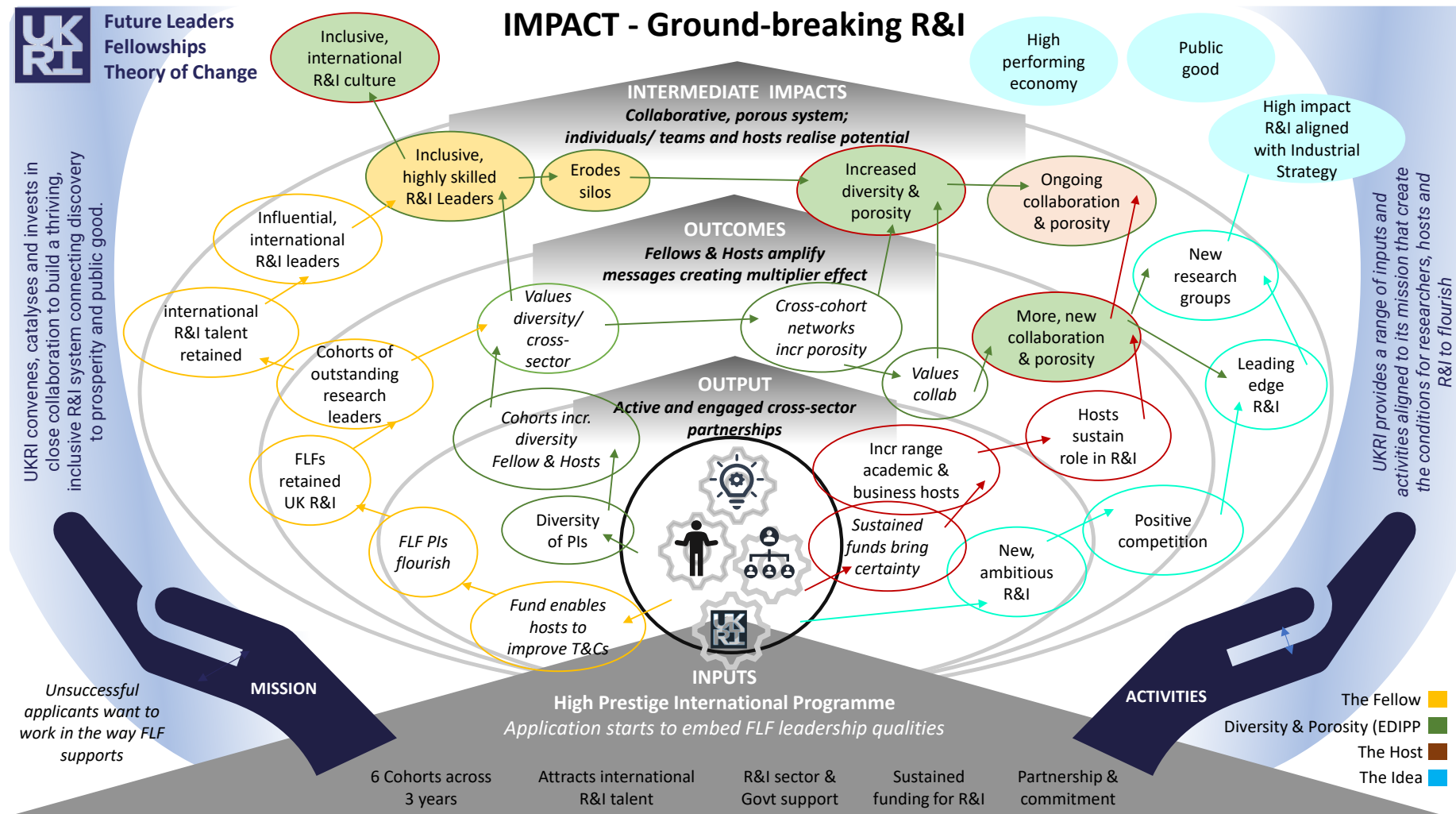
*Extracted from the logic model*

In moving from the logic model to a TOC, there were shifts from the business case (see Figure 2-1). These largely focused on the inputs and activities – which at the stage of making the business case were concerned with inputs and activities scheme design and to gain funding, rather than to the programme itself and how it would support fellows and hosts to achieve outcomes. In contrast, the outputs, outcomes and impacts from the logic model were retained, added to and elaborated. In this latter category are those activities, outputs and outcomes indicated by green shading in the business case logic model which concern the ‘researcher and innovator’ which we refer to as fellows throughout this report; and those with pink shading which encompass outputs, outcomes and impacts from the research and/or innovation (R&I) itself.

To move to a TOC, the aim was to identify the mechanisms through which the FLF Programme is intended work, the effects and impacts it is expected to achieve, and link these through one or more causal pathways. The FLF is a complex programme with multiple aims for impact. As such, there are multiple pathways and mechanisms to achieving the planned impacts. The TOC diagram seeks to capture some of this complexity.

Some mechanisms are more clearly articulated than others, indicating variability in the feasibility of testing different aspects in the evaluation. An example of this is the ambition to affect leadership or cultural change, where measures are not yet fully specified. Part of testing the theory of change in the evaluation may therefore also require an investigation of the specificities of some mechanisms (e.g. providing evidence for the nature of cultural change that is emerging and how this is leading towards final impacts), rather than simply tracking the causal pathway that is indicated.

Figure 2-2: The TOC diagram



## Key to the TOC

The TOC should be read bottom up. *Italics indicate mechanisms*. The lefthand of the diagram focuses on the ‘people and leadership’ components of FLF, while the righthand side centres on the host organisations and the R&I ideas. UKRI’s role throughout the programme is illustrated through the blue elements.

**Figure 2-3: The TOC key**



The project-partnership engine for change. The key components of this are the Fellow, the host organisation and the R&I idea. UKRI plays an underpinning role providing funding and the support that begins to generate the programme effect. The programme aims to support 550 of these engines over the six waves of intakes



The researcher/ Principal Investigator causal pathway



The R&I ‘idea’ causal pathway



The EDIP causal pathway



The host causal pathway

Source: IES, 2021

## 2.4 Inputs and activities

The FLF Programme is intended to be a fellowship programme that is held in high esteem within the research and innovation (R&I) community in the UK and internationally. Assumptions surrounding application include that gaining funding and support through the programme will have caché and be desirable to individuals and hosts. Gaining a Future Leaders Fellowship will be prestigious due to the competitive and rigorous selection process (outlined in Chapter 3), as well as length and high monetary value of awards. Among the key inputs to the programme are the applicants, the hosts who have agreed to support them, as well as the funding UKRI supplies. Key activities include the application process and the support offered by UKRI and host organisations.

To realise this prestigious fellowship programme, the FLF programme design includes inputs and activities in addition the funding itself. The terms of the application and the process of applying for a Fellowship should help to build the attributes and focus that the programme seeks to instil in individuals and hosts. This includes wanting to push new boundaries in R&I, with many fellows doing this through working collaboratively based on multi-disciplinary and cross-sector approaches, and as part of partnerships.

Those who do not gain a Fellowship are expected nonetheless to want to continue to work in the ways their application set-out which will include some who will take forward

multidisciplinary approaches, working collaboratively cross-sector, and working in settings that go beyond the traditional, research intensive universities and centres, to deliver ground-breaking R&I. There is an underlying *assumption that unsuccessful applicants can develop their idea without the FLF funding package i.e. their host organisations will continue their support, although without the same level of funding and support they cannot achieve as much.*

The FLF programme has a strong focus on *porosity* – by which stakeholders mean barriers to collaboration beyond the boundaries of discipline and sector are minimised. Within the FLF Programme, UKRI aims to create the conditions for researchers, hosts and ideas to flourish aligned to its mission. Its mission is to convene, catalyse and invest in close collaboration with others to build a thriving, inclusive research and innovation system that connects discovery to prosperity and public good.

For example, UKRI coordinates the fund on behalf of UK Government. The fund brings together support from the government with the R&I ‘sector’; it ensures there is sustained funding for R&I for the fellows’ career pathways; and it establishes partnership and commitment amongst national stakeholders to deliver the R&I agenda. The assumption underpinning this is that *a virtuous circle is created which has stable conditions for R&I, giving certainty and thereby building confidence in those taking forward ambitious programmes of R&I.*

The Programme mobilises diverse R&I talent, along disciplinary, business and academic lines, through supporting projects located in a range of research councils, that are hosted in business and academic organisations including and beyond traditional, research-intensive organisations. The programme application process removes barriers to this diversity by not specifying disciplinary boundaries and by being open to different interpretations of what is meant by early career.

A key part of UKRI’s activities is delivering a *dynamic support programme* to fellows, that is responsive to their needs and promotes collaboration. This package of support and development from UKRI aims to help fellows maximise the benefits of the programme and develop their R&I leadership skills aligned to the programme’s key goals.

This is a cohort-based programme, and the development programme will build incrementally across cohorts and respond to differing needs. *The assumption is that this leads to a dynamic, development programme that should respond to the needs of each cohort.* The programme will embed skill development and, in order to support ambitions for collaboration and positive competition (competing in such a way that it brings out the best in people and ideas, through being challenging and tapping into potential, while maintaining support and eroding unnecessary silos<sup>2</sup>), it will embed networking opportunities for each cohort and between cohorts.

---

<sup>2</sup> While stakeholders indicated positive competition would emerge, they did not define precisely what they intended by this. The UKRI requested the research team supply a definition which is now to be signed off by the UKRI team.

The host organisations bring structural and other support to fellows, with aims that this will help to accelerate their R&I projects and research leadership role within UK and internationally. The fund should be an engine for partnership with each FLF project embedding a partnership between host, individual, idea and UKRI.

## 2.5 Outputs

The programme and funding should lead to new R&I endeavours involving new, more diverse UK and international talent and host organisations across business and university communities. The programme allows for cross-discipline and cross-sector research partnerships leading to new knowledge and products through maximising the benefits of collaboration and porosity.

The cohort and application process effects (mechanisms) should work in an additive way, continuing to mobilise diverse talent and more hosts into R&I across the years that the programme operates. It is envisaged that FLF participants and hosts will value the gains that can be made from collaboration and porosity (i.e. movement and interactions across sectors and disciplines) and wish to work further in these collaborative contexts, influencing others to value working in the same way. Through this, the effect extends beyond FLF participants and hosts to others in the wider R&I system (e.g. extending out through participants' research teams, departments and cohort interactions).

## 2.6 Outcomes and intermediate impacts

The outcomes and intermediate impacts arise before the achievement of major impacts – which focus on the economy, society, knowledge and aligned to the industrial strategy.

Returning to the business case logic model shown in Figure 2-1, the outcomes and intermediate impacts refer to the point in the logic model where the fellows' outcomes are seen as a precursor to outcomes and impacts achieved by the R&I ideas. The workshop consultations with delivery phase stakeholders showed this was an area where a shift had occurred between design and delivery. This led to the elaboration of planned impacts which related to effects on culture and leadership, as well as a view that fellows' impacts played a more equal role alongside those of the R&I.

The programme is aiming to create a strong cohort of outstanding R&I leaders, who produce positive change in the R&I culture aligned to FLF aims for diversity and porosity, for example through their design and support for research teams, and through their support for one another, as well as through significant R&I outputs in business, society and academia. The culture change that UKRI envisages covers how teams are formed by fellows, if applicable, and the nature of relationships with fellows – the programme will promote good practice on these dimensions. The intention also is that the people involved in FLF will flourish and have an impact on their teams and colleagues in how they understand positive R&I culture. Though this, FLF aims to provide the opportunity to drive culture change.



Fellows are intended to be influential, international researchers, who, building on their FLF experience, go onto embed inter and multidisciplinary working, and/or interactions across business and academia (porosity) in their future projects. They should influence the R&I system, with indicators being that they are invited to take part in grant committees, assessment panels and so on, as well as bidding for research. They are intended to drive R&I within their organisations and the sector more widely. Building on the training and networking opportunities facilitated through the FLF programme, they should multiply the culture of inclusive leadership of R&I and/or teams, and should be conduits for culture change, helping new, diverse UK and international R&I talent to flourish and succeed in R&I, acting as role models.

Fellows and their host organisations should engage in positive competition as well as continue to collaborate developing new partnerships and new ideas.

The relationships between fellows and hosts are intended to lead to important outcomes. The programme is encouraging engagement of a greater diversity of business and academic organisations in R&I through the funding model that expressly focuses on this. The outcome should be their sustained role in R&I, demonstrated through funding and leading ground-breaking R&I in the UK. This would be an amplifying effect from the programme.

## 2.7 Major impacts

The programme is intended to contribute to the UKRI vision for an outstanding R&I system in the UK that gives everyone the opportunity to contribute and to benefit, enriching lives locally, nationally and internationally. As a home for ground-breaking R&I, the FLF should support the UK to evolve a culture that stimulates, attracts, and fosters R&I ideas and talent. Through this, the UK should see major impacts on in its economy (specifically, an increase in R&I-led growth), as well as societal impacts (increases in high-quality jobs and wages and benefits from new products and services), on knowledge (with national capacity for high-quality R&I) and aligned with the industrial strategy (namely, high quality and impactful R&I in areas aligned with the industrial strategy).

A further impact from the programme should be the establishment of a more inclusive R&I culture, visible first through the fellows, their research teams (if applicable) and collaborators. The programme should also bring about change within the wider R&I landscape, increasing the range of business and academic host organisations, supporting partnership between host organisations, as well as influencing how they sustain R&I and talented researchers – which will support the impact to emerge in respect of high-quality jobs and wages. *An underlying assumption is that increased diversity (at host, researcher, project and partnership levels) increases high quality R&I.*

The creation of inclusive, influential research leaders through FLF is an intermediate impact on the pathway to these overall programme impacts.

These inputs, outputs, outcomes, intermediate impacts and wider impacts are summarised in Table 2-2 at the end of the chapter.

## 2.8 Agents for change and the causal pathway

This section provides a summary of how the FLF will effect the changes necessary to achieve the impacts it has set out. It explores this from two perspectives: the agents for change (the people and things that will make this change take place), and the mechanisms for change through which they will operate.

The programme is complex in that it expects impact to be achieved for several groups who are also the 'agents for change'. These are:

- The fellows – interacting with their host, partnership and team (if relevant), and the development programme and cohort network.
- The hosts – providing support to fellows, building on their achievements to develop practice within/beyond their organisation, building new networks that continue to engage and deliver ground-breaking R&I.
- The R&I ideas – which aim to bring about economic and societal change, new knowledge and/or new products.

The causal pathways traverse a ripple created by the Programme (a causal pathway in itself), with the partnership between host, fellow, idea and UKRI being an engine for partnership, diversity and innovation. The ripples (which may be understood as outputs, outcomes and intermediate impacts) are:

- Cross-sector, multi-disciplinary partnerships. These are the engine for change and represent the set-up each fellow and host is part of. The TOC illustrates only one engine though there are multiples.
- Individuals and hosts amplifying and multiplying the effects of the Programme ensuring impact extends into the R&I community and system. Capability, opportunity and motivation combine to lead to behaviours that will underpin delivery of these impacts.
- Collaborative, porous system; individuals/ teams and hosts realise their potential. This will be illustrated by the formation of new research collaborations, a supportive environment for researchers to grow, collaborate and compete, as well as new products, knowledge and services emerging from new collaborative projects.

Through these causal pathways the final impacts – which are expected on the economy and society will be achieved. The pace of change associated with each mechanism and causal pathway is likely to vary; the evaluation could supply insights into the temporal order of different outputs and outcomes, and thereby different causal pathways to impact.



## Mechanisms for change

**Table 2-1: FLF Theory of Change Mechanisms**

Mechanism	Action/ outcome	Position in TOC	Assumptions
The flourish mechanism	Fellows 'flourish' in the programme and work in such ways that their teams/ partners flourish too – contributing to an increasingly encouraging and inclusive research culture.	Output	Fellows operate within host and partnership structures that allow and support this; FLF programme support encourages inclusive and supportive team working. The FLF programme brings the Fellows together, creating spaces for and facilitating interactions, networking and knowledge/ideas exchange.
The multiplier mechanism	Fellows and hosts multiply a culture of good leadership, promoted through FLF networking and training support, and are conduits for culture change, helping others to blossom and training them to succeed, and act as role models.	Outcome	This builds on the flourish mechanism
The diversity and porosity mechanism	Increased diversity (at host, researcher, project and partnership levels) and porosity between disciplines and sectors increases innovation –and thereby impact.	Intermediate impact	Diversity – via networking among fellows from different subject areas and hosts, and via inter and multidisciplinary projects, partnerships and teams – increases innovation, and then innovation increases impact
The cohort mechanism	The FLF accepts two new intakes per year across three years, delivering 550 Fellowships. The process prioritises multidisciplinary projects, and partnerships and more diverse people (across disciplines, hosts and projects) helping to build diversity in the research community	Output	The fund attracts and funds the people, hosts and partnerships that deliver increased diversity, collaboration and porosity needed in the wider system
The application process mechanism	Applying for a Fellowship helps to build the focus the programme seeks to instil in individuals and hosts - wanting to push new boundaries through working collaboratively based on multi-disciplinary approaches, as part of partnerships.	Input	Researchers and hosts understand the unique aims of FLF
The unsuccessful applicant mechanism	Those who do not gain a Fellowship display Fellow attributes e.g. multidisciplinary approaches, working collaboratively and in new partnerships	Output	Unsuccessful applicants persevere with bringing new ambitious R&I ideas, that are built on collaborative working.

The bringing certainty mechanism	The fund ensures sustained funding for the fellows throughout their career path creating stable conditions for their R&I, giving certainty and confidence to PIs and hosts to take forward ambitious R&I projects. The FLF requires hosts and Fellows to commit to R&I for an extensive period. For Fellows, this certainty frees up resource to fully focus on developing leadership skills.	Output	FLF provides initial funding for four years with an option for a further three years funding after this, and hosts commit to longer-term contracts for Fellows at the end of the fellowship <sup>3</sup> .
The host employment contract mechanism	Hosts commit to providing an open-ended UK-based independent R&I position either during or on completion of the fellowship	Output	FLF supports the creation of longer term contracts for research in the sector. Hosts see the value of retaining R&I talent in open ended contracts

Source: IES

## 2.9 From theory of change to evaluation

The Theory of Change provides a schematic of the inputs and activities that FLF will deliver and how these are expected to lead to the planned outcomes impacts. This has been used to guide the evaluation design, by identifying the key outcomes to be measured in an impact study and the causal pathways to be explored through process study and theory-based studies. The major outcomes are retained from the business case logic model and will form the key impact measures for the FLF.

The process and theory-based studies will make clear the underlying causal pathways that result in the pattern of outcomes observed from FLF. These will unpick this in detail, testing whether the hypothesised linkages between the activities and outcome patterns hold in practice and identifying the contextual factors that need to be in place for these relationships to operate (see Chapter 7). Realist evaluations refer to this as the development and testing of *context-mechanism-outcome pattern configurations* (CMOCs) – which comprise models for how interventions activate *mechanisms* amongst whom and in what *contexts*, to bring about the desired *outcomes* (Pawson and Tilley, 2004). The goal of the process study is thus to develop plausible models to explain the pattern of outcomes found in the impact study. The TOC provides the starting point for generating hypothetical CMOCs, which will be further developed, tested and refined through the process study. In turn, they will be used to elaborate and refine the TOC to indicate how and for whom the programme worked.

<sup>3</sup> This latter clause was suggested by UKRI and has yet to be reviewed for agreement across their team.

**Table 2-2: Overview from TOC of inputs through to impacts**

Level	Ambitions and assumptions
Inputs and activities	<ul style="list-style-type: none"> <li>- Strong and diverse R&amp;I talent (removing barriers around disciplines and early career interpretations)</li> <li>- Host commitment</li> <li>- Sustained and high value funding</li> <li>- Rigorous and competitive application process</li> <li>- Dynamic and responsive support programme</li> <li>- Encouraging collaboration, networking and positive competition</li> </ul> <p><i>Assumptions: application process selects the right people who can deliver diversity, collaboration and porosity; stable conditions for R&amp;I gives certainty and therefore confidence to take forward ambitious programmes of R&amp;I; UKRI with support from hosts creates conditions for fellows to flourish and collaborate; collaboration and across discipline and sectors erodes silos and enables ground-breaking R&amp;I; diversity increases innovation.</i></p>
Outputs	<ul style="list-style-type: none"> <li>- New R&amp;I endeavours</li> <li>- Greater volume and diversity of research talent</li> <li>- More partnerships within and across sectors and disciplines</li> <li>- New knowledge and products</li> </ul> <p><i>Assumptions: fellows and hosts continue to influence, attract and mobilise new researchers from wider backgrounds and in wider settings; collaboration is valued and leads to further collaboration and partnership.</i></p>
Behaviours	<ul style="list-style-type: none"> <li>- Push new boundaries in R&amp;I</li> <li>- Collaboration</li> <li>- Multidisciplinary and cross-sector approaches</li> <li>- Leadership and inclusion</li> </ul> <p><i>Assumptions: Application process as well as programme builds desired attributes and focus, unsuccessful applicants will continue to develop, programme embeds skill development</i></p>
Outcomes and intermediate impacts	<ul style="list-style-type: none"> <li>- Strong cohort of outstanding R&amp;I leaders who drive collaboration</li> <li>- Good practice in developing teams and relationships</li> <li>- A more equal, diverse and inclusive research and innovation workforce</li> <li>- Greater availability of R&amp;I careers within new groups</li> <li>- Greater diversity of organisations in R&amp;I</li> <li>- Positive change in R&amp;I culture – EDI, collaboration, multidisciplinary working</li> <li>- Culture that stimulates, attracts and fosters R&amp;I ideas and talent</li> <li>- High quality and impactful R&amp;I strategic areas</li> <li>- Increased impact of R&amp;I through collaboration and problem-solving with business</li> </ul>
Major impacts	<ul style="list-style-type: none"> <li>- Ground-breaking cross-disciplinary R&amp;I</li> <li>- UK maintains world-leading research base post-Brexit</li> <li>- New investment into R&amp;I from outside government.</li> <li>- R&amp;I that enriches lives locally, nationally and internationally through high quality jobs, wages, products and services</li> </ul>

## 3 What is the reality of FLF?

This chapter positions FLF within the wider landscape of fellowship, early career researcher development and innovation programmes highlighting similarities and differences; explains the process used for selection of fellows and the implications this has; explores the make-up of applicants and fellows; and considers the implications of the reality of FLF for planning a full evaluation of the programme.

### Key findings for an evaluation design for FLF:

- Mapping fellowship and similar programmes finds FLF's lengthy support period, large funding amount and size of cohort coupled with its broad disciplinary focus and eligibility to researchers in business settings sets it apart from many fellowship schemes. This makes benchmarking results against other schemes challenging. However, the characteristics of FLF fellows largely follows those of other large early career researcher development programmes: virtually all fellows have a PhD (many gaining their PhD six to 10 years prior to their application), many fellows are in their early 30s. This means comparison to other early career researcher programmes with more rigid criteria could be considered. The potential options for comparisons are Sir Henry Dale Fellowships and European Research Council Starting Grants but there still remain some critical differences. Awardees of these schemes could potentially be used as a comparator group for FLF fellows but would be limited to those in academia and/or in STEM disciplines.
- The application and application assessment process creates data that can be used to establish a baseline for fellows. These data could also be used for a potential comparison group of applicants – as a baseline and to support creation of suitable matches - but there are uncertainties around permissions to process these data for applicants. Using sift and interview scores will be key to creating comparison or counterfactual groups.
- The application assessment process may work against cross disciplinary submissions, there is no standardised threshold to meet to be offered a fellowship, and the assessment scores differ across disciplines and cohorts. These create challenges for drawing comparison groups across the whole programme. A further challenge is the heavy skew in successful applications towards academic hosts (91%) and STEM disciplines (79%) despite FLF ambitions to be neither discipline nor sector specific. Again, this creates challenges for drawing programme level comparison groups and also difficulties defining suitable metrics. It may therefore be appropriate to segment the fellows and consider specific evaluation approaches and metrics for each. Four segments are proposed: STEM academic fellows, AHSS academic fellows, SME fellows and large business fellows.

### 3.1 How different is the FLF programme?

The FLF sits within a large and diverse landscape of innovator and early career researcher (ECR) schemes and fellowships many of which are medium and small-scale. Fellowships

differ from research grants, in that they are designed to support the development of the individual and their career, whereas a grant is focused on the research project and the team of people working on it. However, some fellowships are designed to give early career researchers, individuals who have not yet achieved research independence – not yet leading a project and a research team – to lead research in a Principal Investigator capacity.

Understanding the scope, nature (eligibility and support), and goals of this wider set of schemes and programmes is important for considering an evaluation of FLF as:

- a. Other schemes can have the potential to impact upon the same areas as those outlined in the goals and ambitions for FLF and may make attribution of impact of the FLF on the research system challenging.
- b. There may be other schemes that could potentially provide an appropriate benchmark or comparison group for an evaluation of FLF, although stakeholders agree that the FLF programme has a number of characteristics which make it relatively unique in this field more widely beyond other UKRI programmes and so makes it challenging to find contemporary programmes.
- c. Monitoring and evaluation of other schemes can be useful for scoping methods, measures and data sources used, as well as challenges encountered, to see what can be learned for an evaluation of the FLF programme.

### 3.1.1 Similarity of approach

A review of schemes and initiatives (see Appendix B) aimed at developing research capability in the UK, supporting ECRs and developing research leaders identified the following:

- There are many opportunities for funding and support for researchers, particularly those in their early careers. Overall, between 150 and 170 programmes were identified and mapped, of these 120 were specifically aimed at early career researchers. The rest either aimed at all career stages or those more established (mid-career, or independent researcher). Each individual funder often offers an array of different schemes each year with varying aims, eligibility and support provided.
- There are five key categories of funders: UKRI and its Research Councils, Innovate UK, Research England and the various Catapult centres; Trusts, Foundations and charities (many of which are health focused); learned societies; universities (individually and in groups) and research institutes; and the health sector. Funders tend to support researchers in specific (narrow) disciplines but as some funds/programmes are jointly funded by several organisations these can cross categories and can be deliberately designed to do so. For example, the APEX – the Academies in Partnership in Supporting Excellence in Cross-disciplinary - research award scheme brings together various funding bodies. Similarly, many large trusts and charities which fund support for early-career researchers have linked academic or industry partners.
- The size of the FLF programme is much larger than many other programmes. The anticipated total number of FLF fellows is 550; whereas many other fellowship schemes

especially those not UKRI-funded have fewer than 20 fellows per year (usually between one and five). The exception to this is the longstanding Early Career Fellowships scheme funded by the Leverhulme Trust which has had well over 1,500 recipients, around 150 each year, but these fellowships are narrower in scope and involve considerably less financial support than FLF.

- The lengthy period of funding (over a seven-year period) provided by FLF is quite unusual compared to other programmes. Most range from two to five years and are often linked with the academic year. A few Cancer Research fellowships can last six years, whilst some senior chair programmes can last up to 10 years.
- The size of the funding is much larger than the majority of fellowship programmes. Many just cover all or part of the salaries of fellows for a number of years or provide a much smaller sum to cover some costs; although some offer more substantial funds to cover wider research costs which can include additional researchers and/or technical support. Some of the more generous programmes include: David Phillips Fellows (BBSRC, £1m), Career Development Awards (MRC, £1m), Senior Clinical and Non-Clinical Fellowships (MRC, £2m), Wellcome Trust Investigator Awards (£1m); but these tend to be small in number and limited to specific disciplines
- Many of the programmes are based solely in academia. However, FLF is not completely unique in having host institutions both inside and outside of academia, with some not-for-profit institutions hosting academic-funded recipients and other programmes funding roles based in several locations (including in clinical settings) including the New Innovation Scholars Secondment scheme (UKRI, biomedical sciences), Innovations Partnership Scheme Fellowships and Enterprise Fellows (STFC), Future Generations Industrial Fellowships (Ser Cymru), Industrial Fellowships (Royal Commission of the Exhibition of 1851, Royal Academy of Engineering, Royal Society). These tend to be discipline specific, not necessarily focused on ECRs and make a small number of awards. The multi-disciplinary nature of FLF is also relatively distinctive as is the strong cohort focus. It is common that other programmes are restricted to the disciplines/industries associated with the funder particularly those funded by trusts, foundations and charities, and by individuals funding councils.

### 3.1.2 Potential for benchmarking

There are some programmes which share one or two similar features with FLF but only two programmes appear to share several features with FLF: European Research Council (ERC) Starting Grants and Sir Henry Dale Fellowships. These could provide a potential to benchmark FLF and identify the added value but there remain some differences between these programmes and FLF which are likely to constrain the robustness of comparisons that can be drawn.

- ERC Starting Grants are aimed at early career researchers who are ready to work independently, to provide long-term funding to support excellent investigators and their teams to pursue ground-breaking, high quality/high risk research which is expected to lead to advances at the frontiers of knowledge. They are similar in scale to the FLF fellowships, providing up to €1.5m (plus a further €1m to cover start-up costs) for a



period of five years. They are open to researchers of all nationalities, to a wide range of disciplines covering physical sciences, life sciences and social sciences and humanities, and multi/interdisciplinary proposals are encouraged. However, unlike FLF, a PhD is a prerequisite (and applications must be made within two to seven years of completion). Also, although hosts can be private organisations this does not appear to extend beyond private laboratories; and whilst more than 400 awards are made each year just 62 grants in 2020 were to UK hosts. They are highly competitive, and in 2020 there were 3,272 applications for 436 awards (at a cost of €677m), giving a success rate of 13 per cent. The names and host institutions of the successful researchers are posted on the ERC website.

This group of ERC Starting Grant Fellows could form a potential comparison group for the impact evaluation for FLF for fellows operating in academia, although there are still some differences between the two programmes that would limit the robustness of any comparison.

- The Sir Henry Dale Fellowships offered by the Wellcome Trust and The Royal Society have a similar volume, and level of support to that offered by FLF offering up to £1.2m to cover salaries and research expenses, for five years with a potential extension for a further three years. They too are aimed at early career researchers (or as the Wellcome Trust note 'early-independent or intermediate researchers') to build their own independent research programme and team and establish themselves as an independent research leader in their field. However, these awards are restricted to PhD holders with significant postdoctoral research experience (although there are no limits on time since PhD award), who are expected to have already made significant research contributions such as publications, patents, software development, and are starting to lead their own research. They are also limited to health- related scientific research enquiry. The names and hosts of fellows (and year awarded) are published on the Wellcome Trust website.

Sir Henry Dale Fellows could potentially provide an additional comparison group for FLF, for those in academia and STEM disciplines, despite the programme closing to new applicants after 2021.

## 3.2 Evaluating FLF Applicants

### Overview

There are several elements of the assessment process that have implications for an evaluation of the FLF programme, particularly in determining a potential comparison or counterfactual group:

- Applications, applicants and hosts are assessed by a panel of peers (generally three to four individuals) determined by the field of the proposed project. This may serve to silo filter applications into specific fields and work against cross disciplinary/ MIDRI submissions.

- The evaluation of applicants generates a range of data including scores at pre-sift (out of six), sift (out of 30) and interview stages (out of 30). These could be useful in determining comparison/counterfactual groups. Similarly, the initial full application generates a set of information that can act as a baseline (notably the CV and outputs which have a standard format).
- A range of criteria are used to assess candidates and their proposals at each stage, but the scores generated are not attached to each of the criteria; instead, an overarching score is produced. At each stage, the score is generated by a panel after a review of the evidence and discussion to reach consensus. The overarching score means it is not possible to compare successful fellows with unsuccessful individuals on various elements of their proposals, and so it is not possible to create matches based on the assessment criteria.
- There is no set standardised threshold that individuals have to meet to be offered a fellowship. Also, panels (reflecting broad discipline) are likely to (and do) score differently to each other; and panels are likely to (and do) score differently between the applicant cohorts. This creates challenges for drawing comparison groups across the whole programme.

## Explaining the process

Applicants and their hosts to FLF first submit an outline proposal and then a full proposal. Both involve a standard set of documents. The full application includes as a minimum: proposal form (setting out the host details, project title, start date and duration, detailed costings for the first four years, fellow details and details of any Co-investigators, project aims and objectives, project summary, and beneficiaries, costs for additional staff and their CVs, costs for additional resources, project partners and collaborators, and nominated reviewers), workplan, case for support, justification of resources, data management plan, host supporting statement; and the submission of a three-page CV and outputs list. The latter includes outline employment history, current salary, sector relevant outputs eg publications, pre-prints, patents, new products/processes, trade publications.

The guidance for what to submit is updated for each round and some changes have been made over time to respond to the changing context (notably COVID-19) and in response to learning and feedback about the application process. All applications are made through the Joint Electronic Submission System (Je-S) and the MRC hosts the FLF on behalf of UKRI. Applicants can have only one fellowship application under consideration at a time, so cannot apply to any other scheme led by UKRI or its constituent bodies; they can apply to other funders' programmes but cannot hold multiple fellowships (so if successful in obtaining a grant or fellowship they must withdraw their FLF application); and unsuccessful applicants can re-apply to another round but cannot submit an unchanged proposal and must address feedback given during their failed application process.

Applicants to FLF go through a thorough evaluation process that involves several stages and their application is scored at various stages. Peer reviewers and panel members are asked to consider:



*‘people whose work and careers will benefit most from the scale, duration, and flexibility offered by the FLF Scheme.’*

Central UKRI decides on the most appropriate council for each application. The majority of applications fall under one council (98 per cent), although some have other affiliated research councils. The assigned council then asks experts in the field (related to the application) to peer review the application. Where it is cross council, peer reviewers are sought from multiple councils. Each council aims for three-to-four peer reviews per application, by asking for a minimum of 10 reviewers at the outset. There is variation in the number of reviewers who respond to this peer review request, with the range in Rounds 1 to 4 of between one and eight. Three-quarters of applications (74 per cent) meet the aim of three-to-four peer reviews with 11 per cent receiving five or more peer reviews.

#### ■ Initial assessment – pre-sift

Peer reviewers assess proposals under FLFs four assessment criteria<sup>4</sup>:

- Research and innovation excellence
- Applicant and their development
- Impact and strategic relevance
- Research and innovation environment and costs

Reviewers are encouraged to assess whether the added value of the fellowship structure of support is demonstrated by candidates. Business applications also consider whether the novel nature and level of risk is above that a business would usually take (which is seen as desirable).

The proposals are provided with written feedback by the reviewers across eight dimensions, and a single overarching score (from 1 where the proposal is scientifically or technically flawed to 6 where the proposal is very strong and fully meets all the assessment factors) is given by each reviewer. The feedback and scores are incorporated into the sifting stage.

### Box 3: Application review dimensions

- **Declaration of Interest-** whether the reviewer has a conflict of interest with the proposal.
- **Applicant, Training and Development-** considering the applicant’s track record and current research standing; expertise and skill set; ability to carry out the proposed work; training and development plans for themselves and, if applicable, for team members; leadership potential and; proposed placements and collaborations.
- **Programme-** commenting on the importance, competitiveness and impact of the proposed research and innovation.
- **Host Organisation-** whether the host is appropriate for the work proposed; whether there is a good level of commitment from supervisors, mentors and host institution and if opportunities for training and career development are actively supported.

<sup>4</sup> <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-28102020-FLF-Assessment-Criteria.pdf>

- **Resources Requested-** whether funds requested are appropriately justified to deliver the proposed project; whether the proposal demonstrates value for money in terms of the resources requested and whether any animal use is fully justified and conforms to guidelines.
- **Ethics and Data Management-** whether the proposed research/innovation is ethically acceptable in relation to risks to humans, animals or the environment and if applicable, whether the Data Management Plan indicates the applicant has a sound plan for managing the data funded through the award and in the long-term.
- **Relevance to the Aims of this Scheme-** based on the scheme objectives.
- **Reviewer Expertise-** whether the reviewer has provided comments on the whole or part of the application based on their expertise and experience.

#### ■ Sifting stage

This stage involves a sifting meeting where council-selected review moderating panels recommend who is invited for interview. Round 5 was the largest FLF round so far, with 172 Candidates invited to attend interview. This led to 19 interview panels taking place over two days.

Within each panel, three members are assigned to each application. Should all three “pre” scores from the panel members be below 6, the application is not taken forward into the sift. Should at least one panel review member consider the application of merit/have given a “pre” score of 6, it will be considered further. Individual panel member scores are not collected by UKRI. Once the application is considered further the three sifting panel members calibrate the peer review scores and then provide a consensus score out of 10, using the scoring criteria as set out in the table below. This is the same scoring criteria used by the interview panel.

**Table 3-1: FLF scoring criteria (for sifting and interview)**

Overall Score	Usual indicators
10	<p>The proposal is exceptional; it very strongly meets all of the assessment criteria.</p> <p>The proposal is at the leading edge internationally, addresses all of the assessment criteria, and meets them all to an exceptional level. The candidate has demonstrated that the Fellowship will provide exceptional added value.</p> <p>The panel agree that it is difficult to articulate how the proposal could be improved.</p>
9	<p>The proposal is outstanding; it strongly meets all of the assessment criteria.</p> <p>The proposal is at the leading edge internationally, addresses all of the assessment criteria, and meets all of them to an outstanding level. The candidate has demonstrated that the Fellowship will provide outstanding added value.</p>
8	<p>The proposal is excellent; it strongly meets the assessment criteria.</p> <p>The proposal is of a high international standard and addresses the majority of the assessment criteria to a very high level. The candidate has demonstrated that the Fellowship will provide excellent added value.</p>
7	<p>The proposal is very good; it meets the assessment criteria well but with some minor weaknesses/limitations.</p> <p>The proposal is internationally competitive. It has some weaknesses but meets the majority of assessment criteria to a high level. The candidate has demonstrated that the Fellowship will provide very good added value.</p>

6	<p>The proposal is good; it meets the assessment criteria well but with some clear weaknesses/limitations.</p> <p>The proposal has a number of weaknesses but generally meets most of the assessment criteria well. The candidate has demonstrated that the Fellowship will provide good added value.</p>
5	<p>The proposal is adequate; it meets the assessment criteria adequately but with clear weaknesses/limitations.</p> <p>The proposal meets some assessment criteria well, but has weaknesses relating to a number of criteria. The candidate has demonstrated that the Fellowship will provide adequate added value.</p>
4	<p>The proposal is weak; it meets the assessment criteria but with significant weaknesses/limitations.</p> <p>The proposal is potentially of some merit and meets all of the assessment criteria to an adequate level but is not internationally competitive. The candidate has not demonstrated that the Fellowship will provide added value.</p>
3	<p>The proposal is poor; it meets the assessment criteria but has major weaknesses/limitations.</p> <p>The proposal is potentially of some merit, and weakly meets all of the assessment criteria, but is not internationally competitive. The candidate has not demonstrated that the Fellowship will provide added value.</p>
2	<p>The proposal is unsatisfactory; it does not meet one or more of the assessment criteria.</p> <p>The proposal is of limited research/innovation merit or contains significant flaws. The candidate has not demonstrated that the Fellowship will provide added value.</p>
1	<p>The proposal is unsatisfactory; it does not meet any of the assessment criteria.</p> <p>The proposal is of limited research/innovation merit or contains significant flaws. The candidate has not demonstrated that the Fellowship will provide added value.</p>
0	The proposal is recommended for deferral.

*Source: Future Leaders Fellowships – Round 5 Interview Panel guidance*

The consensus score across the three sifting panel members is decided after a discussion. Roving panel members go across all panels to impose consistency in the scoring criteria. In addition, each application is assigned a low, medium or high score alongside the numeric score. These are defined as follows:

- High – at the top of that scoring definition and close to scoring one higher, e.g. “all things being equal this is a seven, but with a few small improvements would have been an eight”.
- Medium – firmly within that score
- Low – meeting the description of the score but only minimally, e.g. “all things being equal this is a seven, but is close to being scored a six”.

This effectively delivers a 30 point sift system (eg Low 6, medium 6, high 6 then low 7, medium 7 and high 7 etc). Graphical representation of scoring in this report is based on this derived scale.

Each application is then banded at sifting stage from A-D, the definitions are set out in the table below. Based on the sift scores, each of the 14 panels all have their own

thresholds for these bands. Therefore, across councils and rounds the score needed to be in band A may differ due to the strength and number of applications.

**Table 3-2: FLF banding descriptions**

Score	Possible action
A	These candidates are judged to have addressed all of the assessment criteria and show evidence of the highest standards and obvious added value from a FLF award. They are a priority for investment.
B	These candidates are judged to have addressed all of the assessment criteria and are of a high standard, they show good evidence of added value from a FLF award. They should be considered for investment but are not prioritised.
C	These candidates adequately address all of the assessment criteria but are not of the highest quality across the criteria. They are the lowest priority to consider for investment.
D	These candidates failed to adequately address one or more of the assessment criteria and are not recommended for investment.

*Source: Future Leaders Fellowships – Round 5 Interview Panel guidance*

#### ■ Interview stage

Prior to the interview, panel members are encouraged to read both the proposal documentation and the information provided by UKRI in the panel agenda which summarises discussions of the proposal held at the sift meeting. This may include any suggested interview questions highlighted as part of the sifting process. UKRI does not require any written comments or scoring to be submitted in advance of the interviews.

Each interview panel contains four-to-five members plus a Chair, each with assigned responsibilities during the interviews. Each panel will interview approximately 10 candidates. The interviews use a set of standard questions to ensure consistency across interviews and panels, and candidates are required to give a short presentation on 'The vision for my Future Leaders Fellowship'. The interview lasts 30 minutes for the candidate. Candidates are scored using the same criteria as the sifting process.

### 3.3 Diversity of FLF applicants and awardees

As noted in Chapter 2, the FLF is deliberately not sector specific, and aims to cover a range of disciplines and support fellows with various backgrounds and in different contexts including academia and business, independent research organisations, research council institutes, Innovate UK catapult centres and other public research organisations.

As such, it is important to understand the extent of diversity within the FLF population (the fellow and applicant pools) across personal and institutional characteristics, as this gives insights into some of the inputs at whole population level, which may have implications for the planned impacts, as well as for a robust impact evaluation to be taken forward.

Descriptive analyses of who applies for and participates in FLF are based on management information (MI) extracted in January 2021 for Rounds 1 to 4 (as these rounds had been completed and awards made).

### 3.3.1 Overview

Despite the ambitions for the programme, the applicant pool and the distribution of awards is not as diverse as may have been anticipated:

- Over the first 4 rounds of applications (there will be 6 rounds in total) 1,553 applications were received and 305 fellows funded. The volume and quality of applications appears to have increased over time and the application process has become more competitive.
- The MI shows that 54 per cent of fellows are female, 46 per cent are aged 31 to 35, 14 per cent are from a minority ethnic group, 5 per cent have a disability, and 47 per cent are non-UK nationals.
  - The programme has improved over time in attracting applications from women, women do well in the application process (having a higher award rate than men) and this is better than found in other programmes. There is a heavy gender bias by discipline which is not unusual.
  - FLF also does well with regards to nationality, attracting applicants and making awards to a wide range of nationalities, and does better in this respect than other programmes particularly recently with the UK's withdrawal of the EU.
  - Although the programme attracts applicants from all age groups, almost half of all awards have been made to those in their early 30s (aged 31-to-35). Very few applicants and fellows report a disability. Both these patterns are common among fellowship programmes aimed at early careers researchers.
  - The award rate amongst those from a minority ethnic group is lower than found for those from a white background, particularly in the Rounds 3 and 4, this also reflects patterns found in other fellowships.
- Awards are heavily concentrated in academia (and among Russell Group universities) and in STEM subjects. Close to four in five fellows (and/or their projects) (79 per cent) are STEM focused, and 91 per cent are hosted by an academic organisation, 5 per cent in business, and 4 per cent by other hosts. There is a differential award rate by discipline (lowest in ESRC and AHRC reviewed applications) and by type of host organisation (lowest in business organisations), This means it is hard to define appropriate metrics across research areas, and makes it difficult to detect impact for fellows based in business due to the smaller sample. This supports the idea of detecting impact for subgroups of the FLF population.
- Less than one per cent of fellows do not have a PhD, 47 per cent had obtained their PhD six to ten years prior to their application, and 4 per cent work part-time study. This suggests that the intention for a wider diversity of Principal Investigators to emerge has not been realised, at least in the early rounds of awards. This may mean that comparison to other early career programmes with more rigid criteria could be considered.

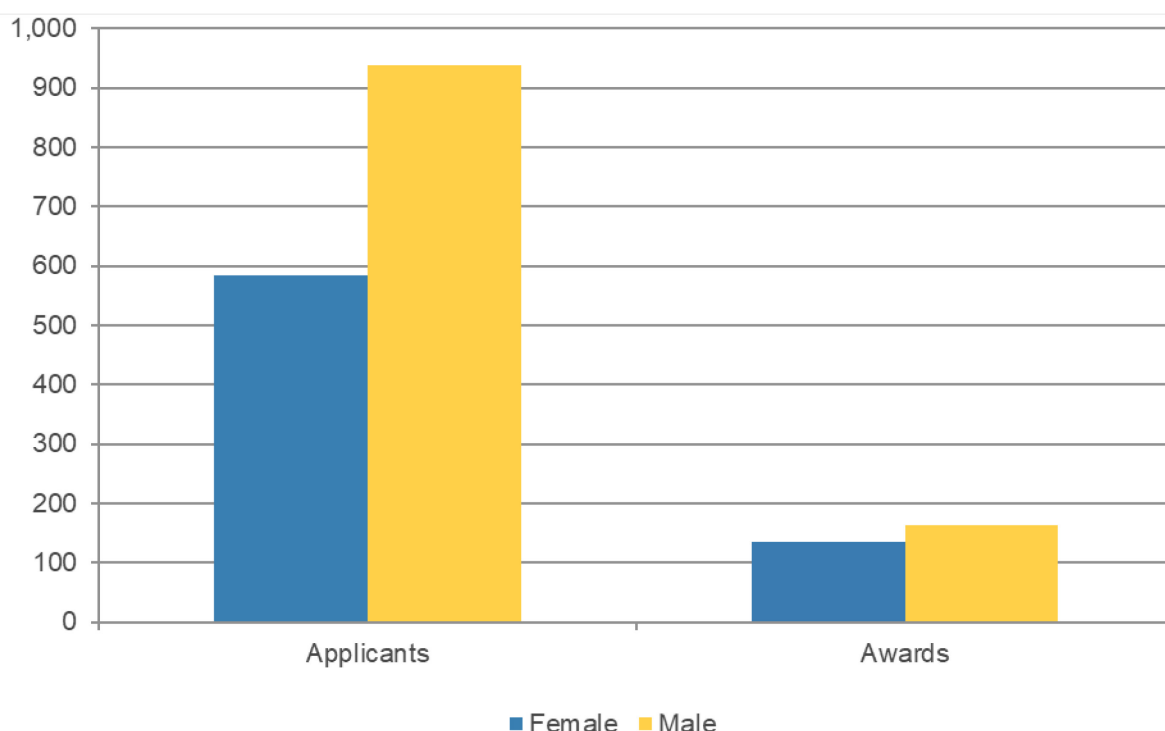
### 3.3.2 Demographics

#### Gender

Figure 3-1 shows that the applicant and award pools for the FLF programme have a very different gender make up. In the first four rounds, the applicant pool was skewed towards males (driven heavily by the number of applications in the field of engineering and physical sciences) but the numbers of male and female awardees were similar (only 26 more awards given to males than females). This application and award disparity is common among UKRI fellowships and is visible in the UKRI diversity data release<sup>5</sup>.

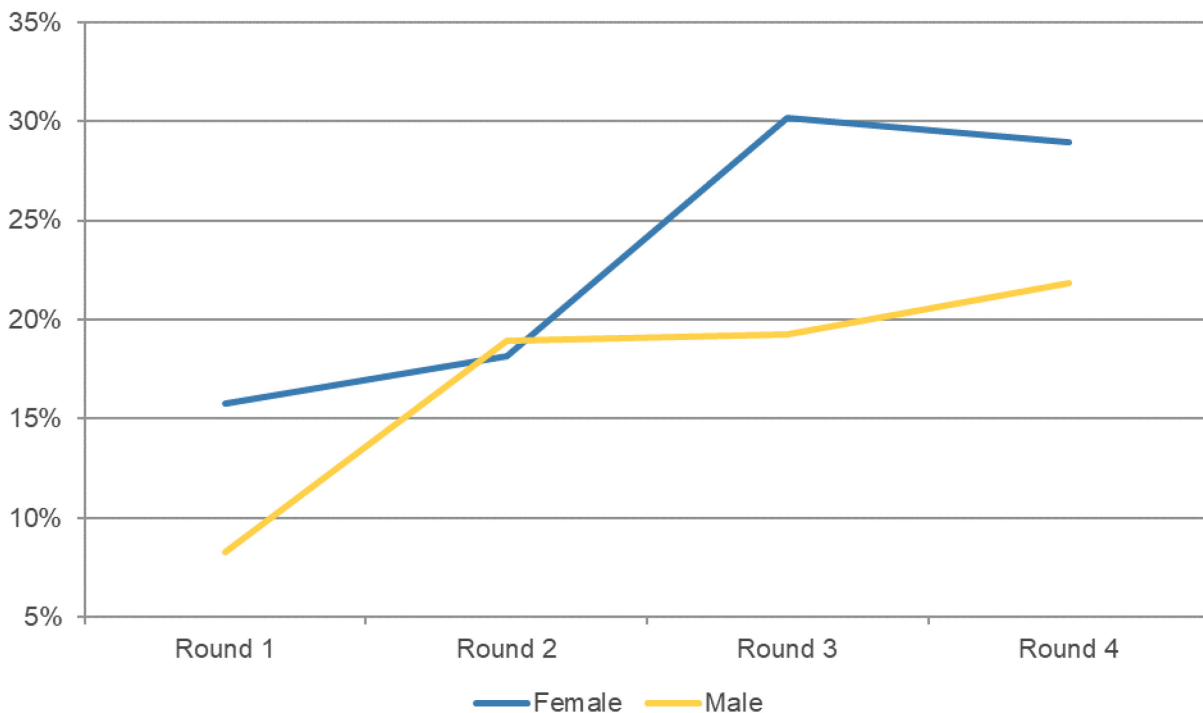
Overall, 38 per cent of applicants were female, and 46 per cent of offers were to females. The award rate to females has risen since the start of the programme from 16 per cent in round one to 30 per cent and 29 per cent respectively in rounds three and four. Figure 3-2 shows that the award rate for males has risen from a low of 8 per cent in round one to 22 per cent in round four.

**Figure 3-1: Applicants and awards by gender**



Source: FLF management information, Rounds 1-4

<sup>5</sup> [www.ukri.org/about-us/equality-diversity-and-inclusion/diversity-data](https://www.ukri.org/about-us/equality-diversity-and-inclusion/diversity-data)

**Figure 3-2: Award rate by round by gender**

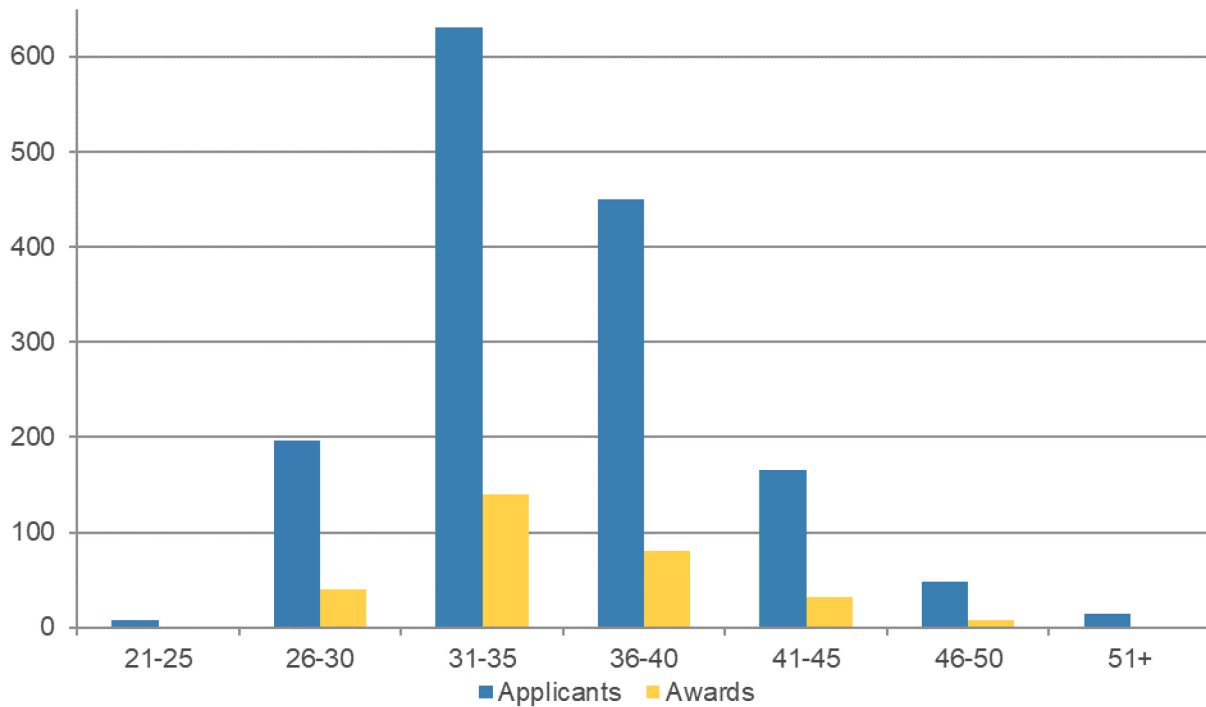
Source: FLF management information, Rounds 1-4

## Age

Figure 3-3 shows that all awards have been given to those aged between 26 and 50 despite a small number (23) applications outside this age range. Almost half (46 per cent) of all awards so far have been given to those in the 31 to 35 age-bracket, the common age-bracket for early-career researchers. This age group also had the highest award rate (at 24 per cent). More than one in four fellows is aged between 36 and 40 (27 per cent), with just three per cent of fellows aged 46 to 50.

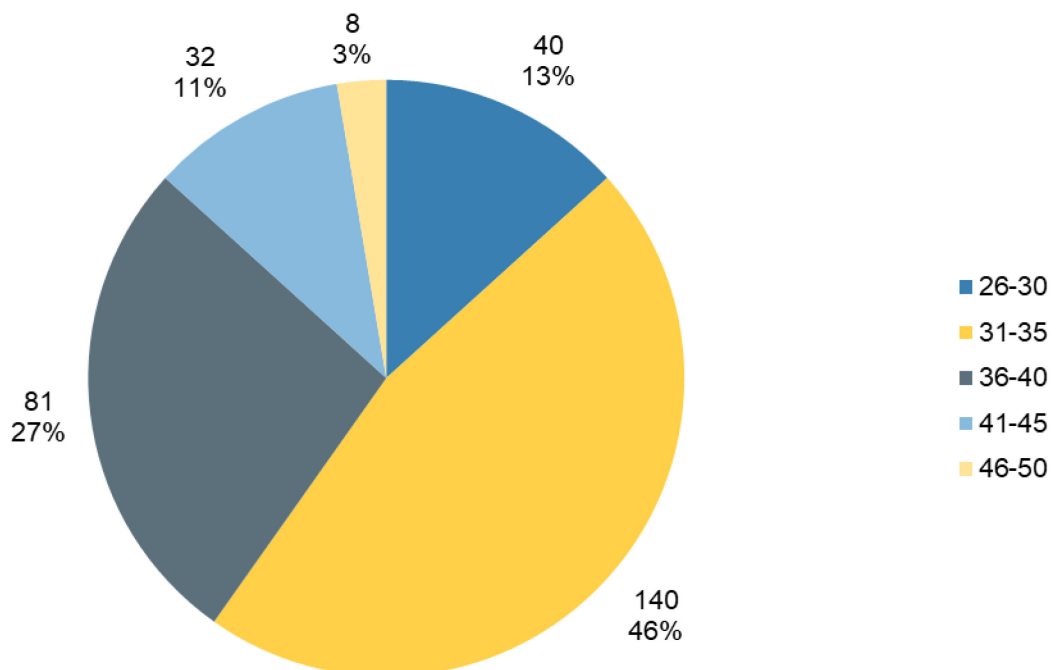


**Figure 3-3: Applicants and awards by age group**



Source: FLF management information, Rounds 1-4

**Figure 3-4: Awards by age group**



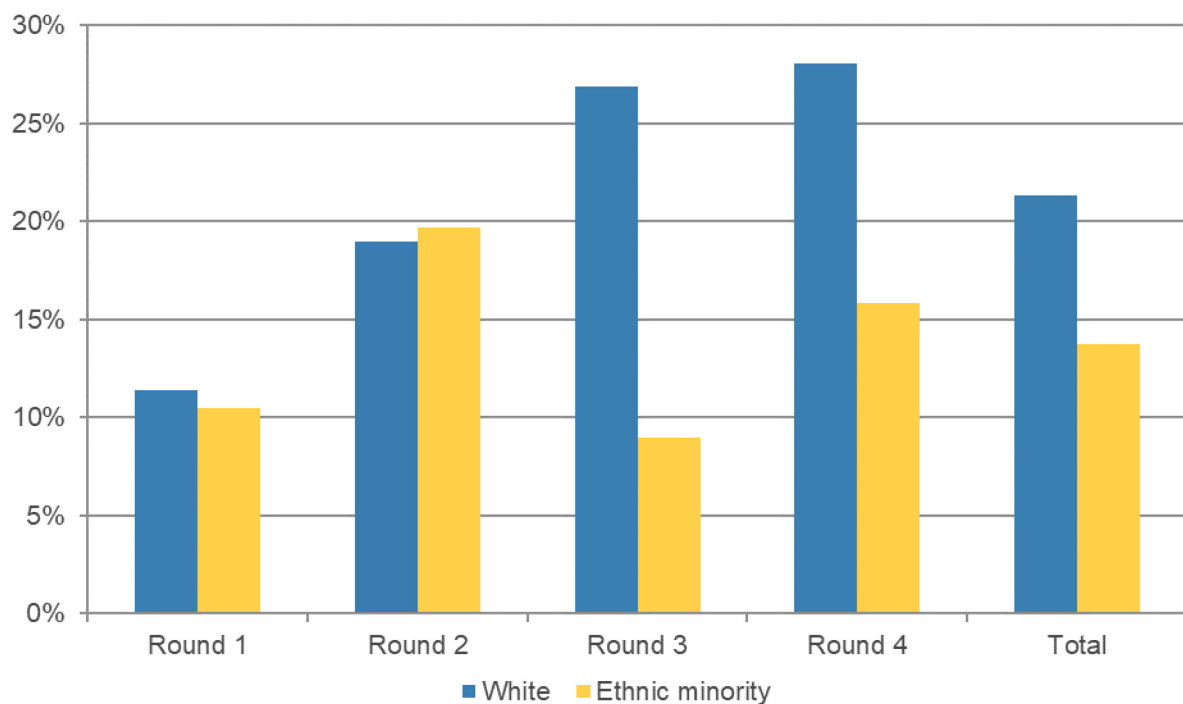
Source: FLF management information, Rounds 1-4

## Ethnicity

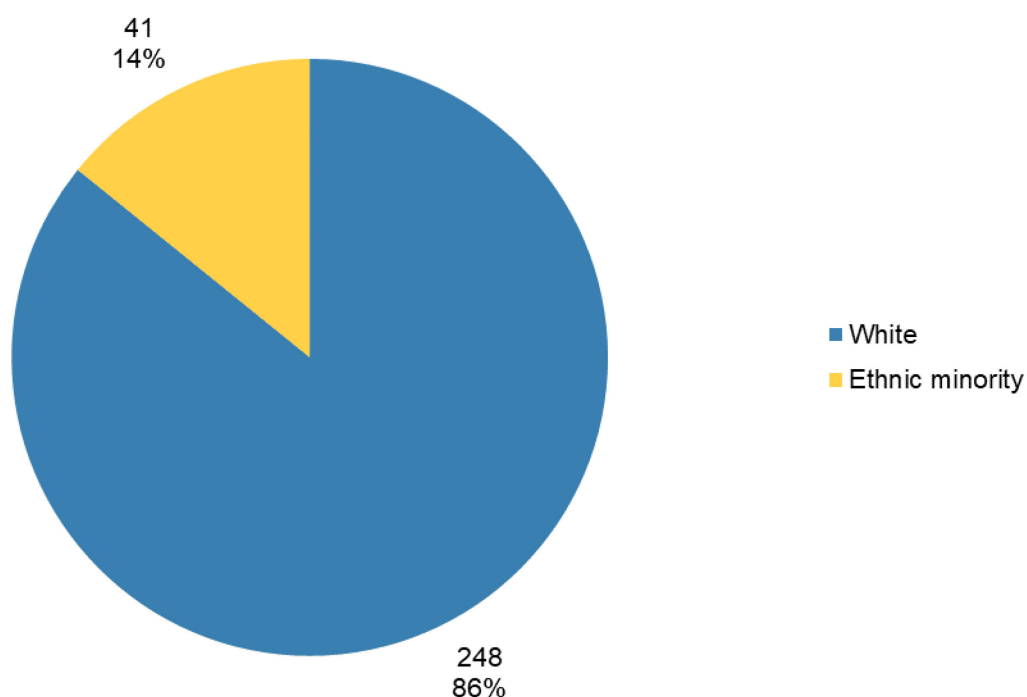
People from minority ethnic groups make up one-in-five FLF applicants (20 per cent), and 14 per cent of all awards. In total, 41 awards have currently been made to someone from a minority ethnic group.

Whilst the award rate was similar in Rounds 1 and 2 between those from a white and minority ethnic background, there were marked differences in Rounds 3 and 4. For example, those with a white background were three times as likely to have received an award in Round 3 than those with a minority ethnic background. The award rates in Round 4 were 28 per cent and 16 per cent respectively, for those from white and ethnic minority backgrounds. Overall, there is an eight-percentage point difference in award rate between those from a minority ethnic vs a white background.

**Figure 3-5: Award rate by round by ethnic background**



Source: FLF management information, Rounds 1-4

**Figure 3-6: Awards by ethnic background**

Source: FLF management information, Rounds 1-4

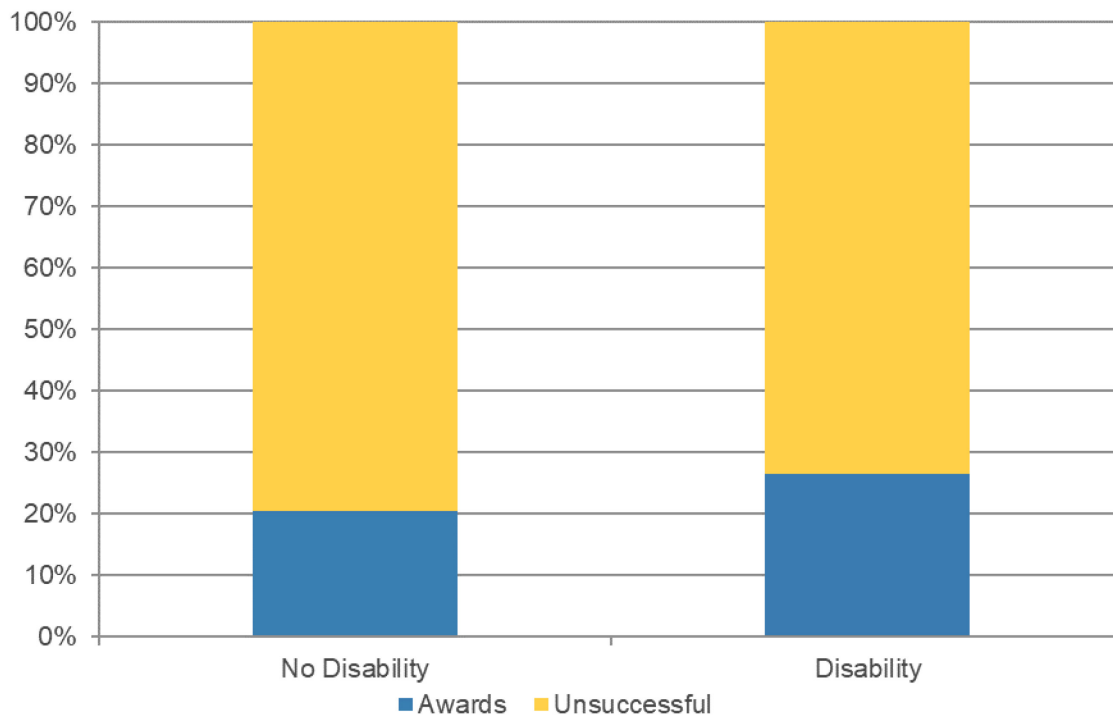
### Disability

In total, three per cent of applicants declared themselves as having a disability<sup>6</sup>. This figure is not too dissimilar to other fellowship programmes, although is much lower than the population as a whole. HESA staff data from 2019/20 show that just below five per cent of academic staff reported having a disability<sup>7</sup>. The award rate is higher among those with a disability (27 per cent) compared to those who are not disabled (20 per cent). This represents a six-percentage point gap however is not statistically significant.

So far there have been 51 applicants who have declared a disability and 13 awards among these candidates.

<sup>6</sup> A further 106 applicants (7%) did not disclose and 6 applicants (<1%) it was unknown whether they had a disability.

<sup>7</sup> HE academic staff by disability and academic employment function 2019/20. See: <https://www.hesa.ac.uk/data-and-analysis/staff/table-5>

**Figure 3-7: Award rate by disability status**

Source: FLF management information, Rounds 1-4

### 3.3.3 Pathways

#### Qualification and experience

It is worth noting that, although the programme is open to individuals who do not hold a PhD, the vast majority of applicants and successful fellows possess a PhD. Across Rounds 1 to 4 (where data is available) approximately 1 per cent (17 out of 1,553) applicants did not have a PhD, and less than 1 per cent of fellows did not have a PhD. The success rate of those applying without a PhD was lower (14%).

The programme also does not have a minimum or maximum time since completing a PhD as an eligibility criterion, and indeed there were examples of applicants and fellows having gained their PhD some considerable time ago. Although there were applicants (n=6) having gained their PhD over 20 years ago, the longest period among fellows was 18 years since completing their PhD. Among successful applicants ie fellows across Rounds 1 to 4, the breakdown is: 9 per cent having completed their PhD within 2 years, 33 per cent between 3 and 5 years previously, 47 per cent (the largest group) between 6 and 10 years previously, and 12 per cent more than 10 years previously. The success rate is fairly consistent regardless of when applicants completed their PhD (20 to 22 per cent), with the exception of those completing within 2 years of making an application to the FLF programme where the success rate was lower (14 per cent).

Another aspect of FLF is its flexibility, particularly in offering support for those wanting to work part-time. The monitoring data for Rounds 1 to 4 shows that approximately 4 per cent

of all applicants were for part-time and similarly 4 per cent of fellows were part-time. There was no difference in the success rate depending on preferred mode.

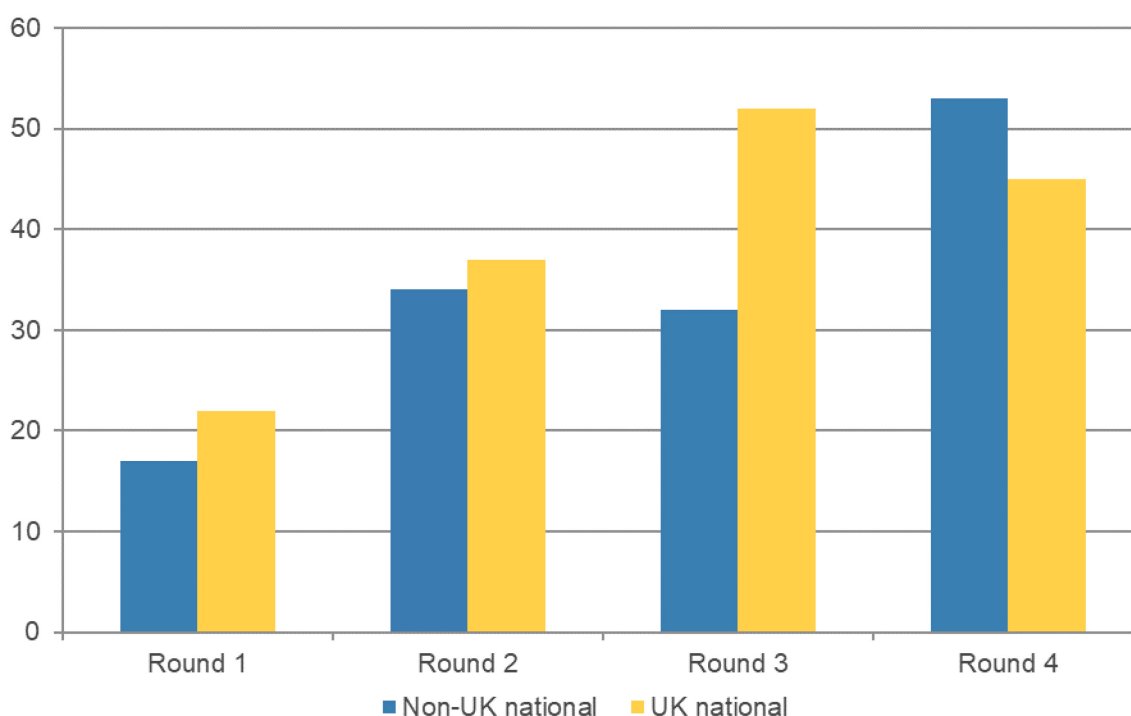
### Nationality

There is a 50/50 split between UK and non-UK national applicants to the FLF scheme. There were higher acceptance rates to the programme among UK national applicants than non-UK nationals for the first three rounds of the programme however only in Round 3 was this statistically significant. This is shown clearly in the Figure 3-8 below. Whilst the number of awards offered to non-UK nationals was higher in Round 4 than those offered to UK nationals it hasn't fully recovered this gap. Currently 47 per cent of fellows are non-UK nationals and 53 per cent are UK nationals.

The applicant pool up to Round 4 was made up of 74 nationalities, whilst the fellow pool is made up of 36 nationalities. There were more than 30 applications from the following nationalities: UK (744); Italian (102); German (74); US (60); Chinese (51); Indian (39); Spanish (39); Greek (36) and Irish (36). There are more than 10 fellows who are UK (156); German (17); Irish (16); US (14) or Italian (11) nationals.

It is expected that the majority of fellows (including non-UK nationals) wish to pursue a research career within the UK, but if individuals move abroad after completing their fellowship they may be challenging to track (eg in UK based secondary data sources or via survey if contact details are not updated).

**Figure 3-8: Award rate by nationality (UK vs. non-UK)**



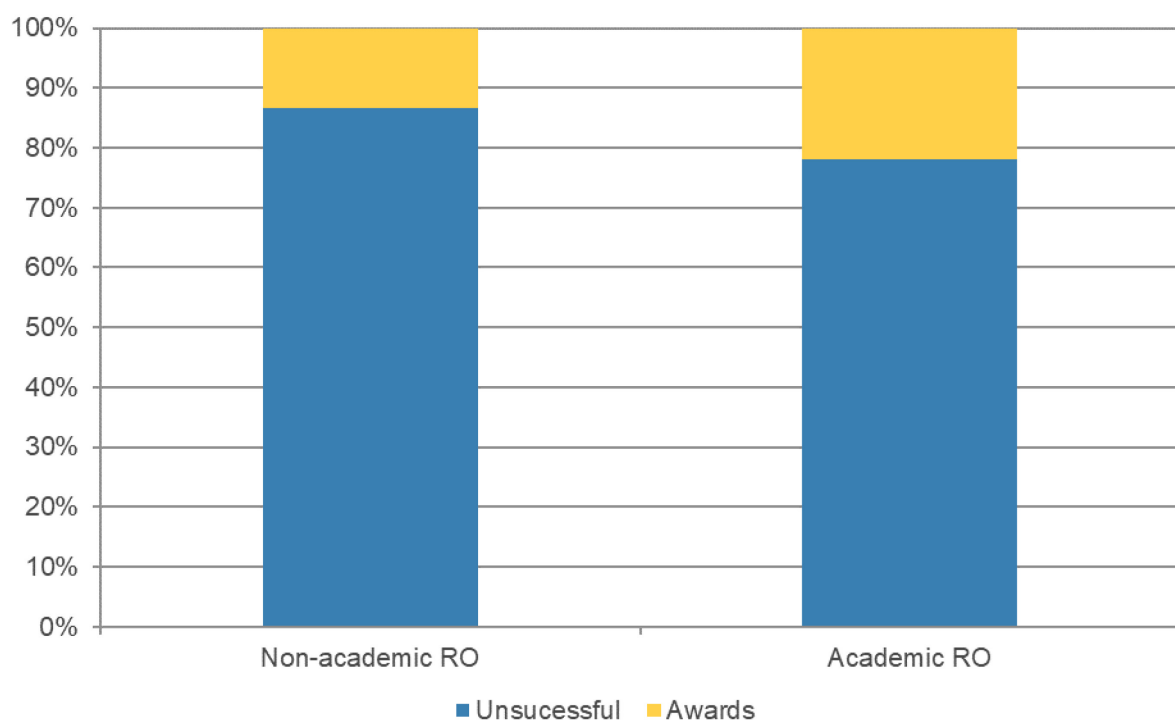
Source: FLF management information, Rounds 1-4

### Host research organisation

The majority of applications were from applicants who would be based at an academic institution with just one in ten (9 per cent) applicants based in business. Other research organisation types include independent research organisations, research council institutes, Innovate UK catapult centres and other public research organisations. Combined these make up five per cent of all applicants.

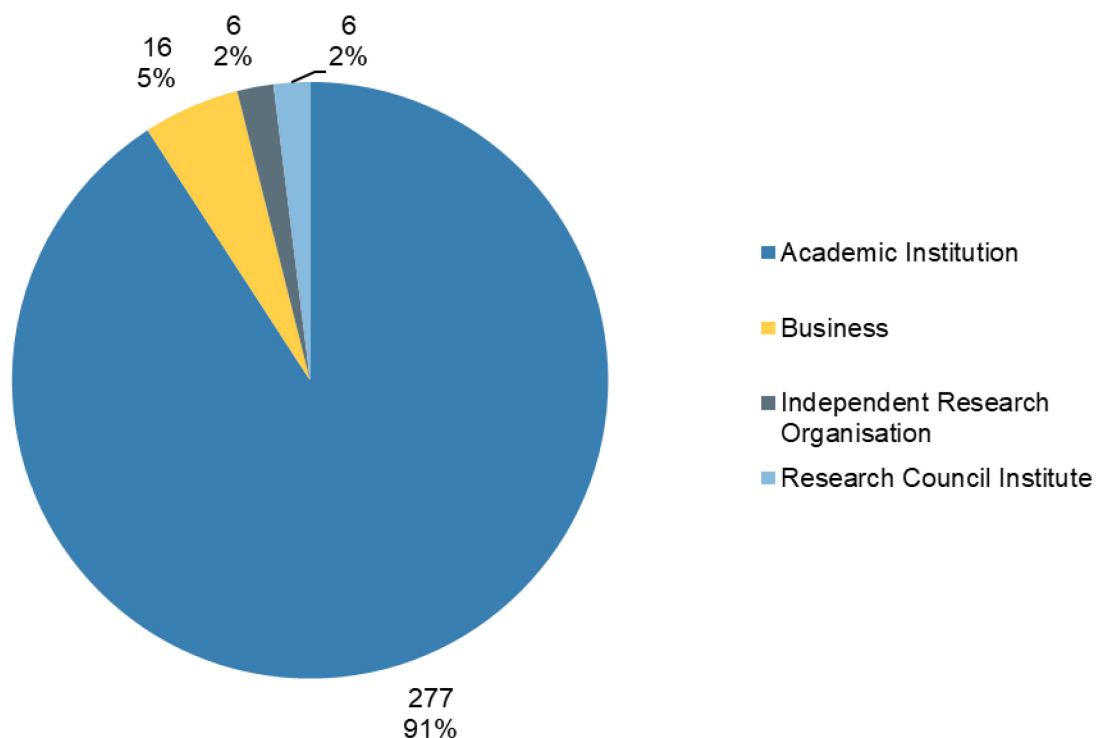
There are differences in likelihood of award based on type of host, with overall a lower award rate outside academia (Figure 3-9). There are indications of greater variability in the quality of applications from business. In over one in ten applications involving a business host, the host withdrew before the sifting stage. This compares to four per cent across all applications. Business applications were more likely to be rejected at the sifting stage (four in five were rejected at this stage), with average sift scores for applications with business hosts of mid-6 compared to mid-7 among academic applications. However, at interview stage there was no real difference in the likelihood of receiving an award between business and academic hosts.

**Figure 3-9: Award rate by institution type**



Source: FLF management information, Rounds 1-4

The lower award rate outside academia combined with the lower number of applications outside academia, means that FLF fellows are largely based in academia: just nine per cent of awards are outside academia. There are sixteen awards based in Business (5 per cent), six based in Independent Research Organisations (2 per cent) and six based in a Research Council Institute (2 per cent).

**Figure 3-10: Awards by RO type**

Source: FLF management information, Rounds 1-4

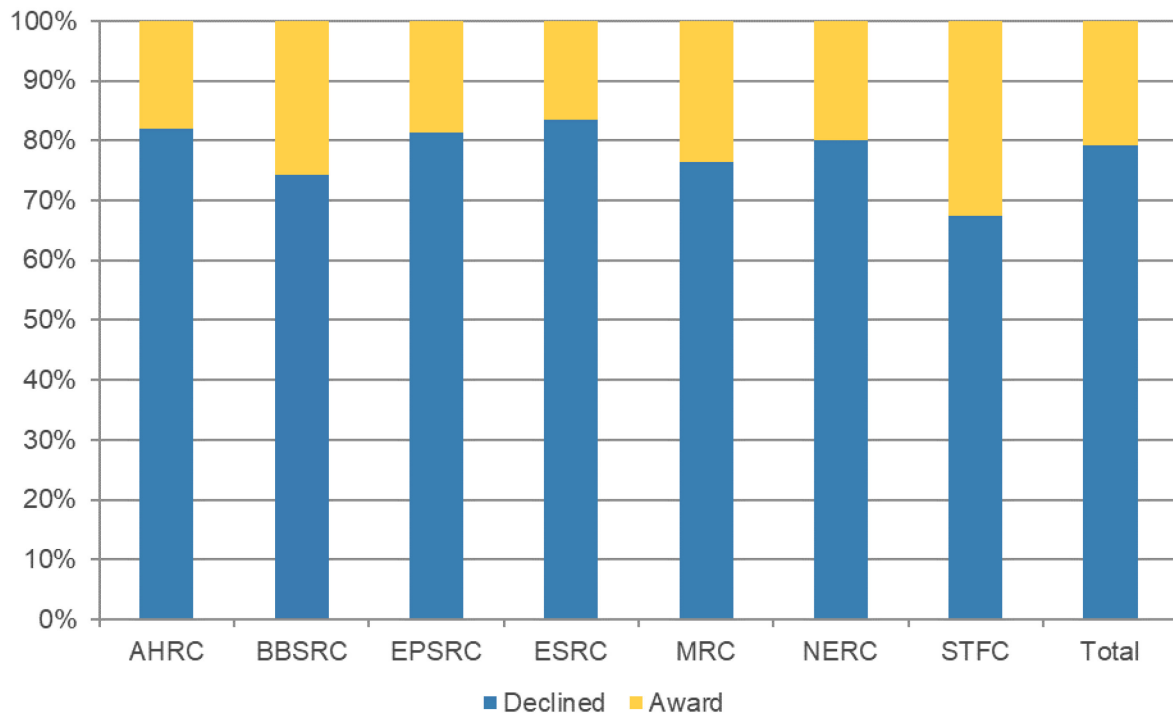
### Research council

Of the applicants who were assigned a research council, the largest number were assigned by the UKRI central team to the EPSRC, nearly one in three applications (473). Both MRC and ESRC were assigned around 250 applications in the first four rounds. STFC had the fewest applications assigned, 98. Despite this the likelihood of award was highest in this council with one in three applicants awarded an FLF fellowship. The other science-based research councils BBSRC and MRC also have higher success rates than average (26 per cent and 24 per cent respectively). ESRC and AHRC have the lowest award rates at 17 and 18 per cent respectively.

Of the current awards, three in ten were assigned to the EPSRC, with one in five assigned to the MRC.

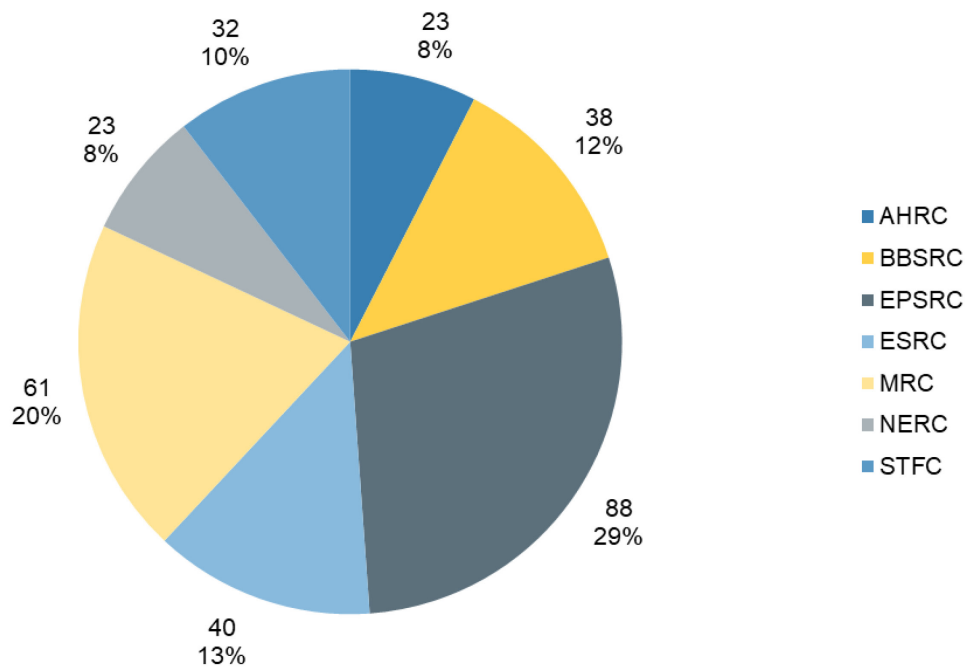


**Figure 3-11: Award rate by UKRI council**



Source: FLF management information, Rounds 1-4

**Figure 3-12: Current awards by UKRI council**



Source: FLF management information, Rounds 1-4

### 3.4 Comparing participation in FLF

It is useful to contextualise FLF in comparison to other multidisciplinary fellowship programmes: to understand who participates in other similar schemes in terms of their characteristics, and the degree to which these individuals are concentrated in particular sectors or disciplines. Two key programmes have been identified for comparison: Royal Society Early Career Fellowships and Wellcome Trust Fellowships. These were chosen as they are among the larger programmes aimed at early career researchers, and also they publish demographic data on their applicants and successful fellows.

Royal Society (The Royal Society, 2020<sup>8</sup>) analysis of their three early career fellowships (Dorothy Hodgkin Fellowship, University Research Fellowship and Sir Henry Dale Fellowship) show that in 2019 across all ECR fellowships, 38 per cent of applicants were female with 39 per cent of the offers were made to females. This represents an upward trend from previous years of the programmes. Overall, 21 per cent of applicants were from individuals from an ethnic minority background in 2019, compared to 16 per cent in 2018. The proportion of offers made to those from an ethnic minority background in 2019 was just 8 per cent, down from 14 per cent in 2018. Very few Royal Society applicants declared having a disability (3 per cent).

Wellcome Trust analysis (Wellcome Trust, 2020<sup>9</sup>) of their grants show that for Early Career and Intermediate Fellowships, women and men both had a 9 per cent award rate. However, analysis shows the proportion of awards given to women is higher than found for the Royal Society: with 46 per cent of awards made to women in 2014/15 and this rose to 51 per cent between 2014/15 and 2018/19. Overall awards on schemes aimed at early and mid-career researchers are more likely to be made to women but there are differences by discipline: awards amongst Science programmes are less likely to be made to women (38 per cent), whereas for Culture & Society schemes the proportion of offers made to women is much higher (59 per cent).

In terms of other characteristics:

- Just two per cent of applicants over the period from 2016 to 19 identified themselves as disabled. The award rate was slightly lower for those with disabilities (13 vs. 15 per cent), but the level is too low to identify a statistically significant difference.
- The mean age of awardees has been broadly consistent since 2016. The average age of Investigator Award recipients was 49 to 50 years; Senior Fellows 43 to 44 years; Intermediate Fellows 37 to 38 years and Early Career Fellows 32 to 34 years. Around 20 per cent of awards are for those aged 50 and older, 40 per cent are to those aged 29 or under (of whom, the vast majority received PhD Studentships and Vacation Scholarships).

<sup>8</sup> [https://royalsociety.org/-/media/policy/topics/diversity-in-science/DES6507\\_Annual-Diversity-Data\\_Report-2019-FINAL.pdf?la=en-GB&hash=1ACEF2C50ED7CB448013118609DF49DA](https://royalsociety.org/-/media/policy/topics/diversity-in-science/DES6507_Annual-Diversity-Data_Report-2019-FINAL.pdf?la=en-GB&hash=1ACEF2C50ED7CB448013118609DF49DA)

<sup>9</sup> <https://cms.wellcome.org/sites/default/files/grant-funding-data-2018-2019.pdf>

- In the 2018/19 funding period, two thirds of applicants were by British nationals. For the Early Career and Intermediate Fellowships there was a fall in applicants from EU nationals in these two programmes from 40 to 30 per cent and 30 to 25 per cent respectively.
- The ethnic make-up of the applicant pool of UK- based applicants mirrors the ethnic profile of the UK as a whole, however the rate at which awards are offered to those with from an ethnic minority background (15 per cent) is lower than the rate awarded to white applicants (18 per cent). This difference is noted across most Wellcome Trust schemes.

On a broad level the demographic profile of FLF does not notably differ to other multidisciplinary fellowships offered by other funders in the UK. FLF has a higher award rate for women, not atypical among other fellowship programmes including those offered by UKRI councils. The large science-based applicant pool drives the fact that more offers have been given to males than females, but the fellow pool is more diverse, for example than those offered Royal Society fellowships. FLF has a similar age profile to other early career researcher-based programmes. Despite attracting a more ethnically diverse applicant pool, the gap in award rate for FLF between those with a white ethnic background and those with an ethnic minority background, leads to a less ethnically diverse pool of fellows. Whilst schemes such as those offered by the Wellcome Trust also experience an award rate differential, it is perhaps not as large as FLFs. Applications from those who declare themselves as having a disability is low, but not atypical to other programmes or the wider academic population.

## 3.5 Other considerations from FLF programme data

### 3.5.1 Defining sub-groups

The analysis of programme monitoring data above shows a number of skews in the FLF cohort with implications for: (a) possible metrics and the potential to apply the same set of metrics across the entire FLF population; and, b) the potential evaluation methods that can be used due to the differing nature of the cohort, and relative size of potential segments.

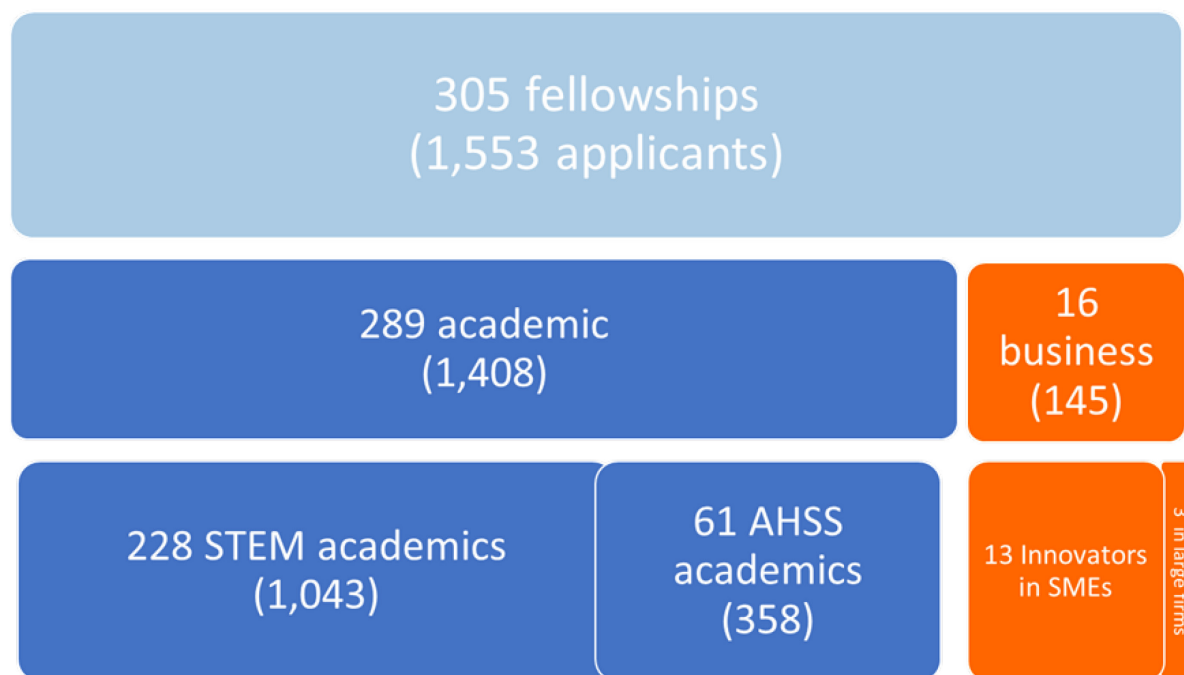
The monitoring data clearly shows that:

1. Take-up of FLF is higher among those from academia than found for business. Due to the small sample of applicants and fellows from business organisations a robust impact evaluation is unlikely to be possible. This is further complicated by considerations of business size, as outcomes and impacts in a business setting will be harder to identify (and attribute) in large organisations. The programme monitoring information for Rounds 1 to 4 shows the majority of business-based (applicants and fellows) are in small and medium size enterprises, with just three in large firms at this stage.
2. Take-up is also higher among scientific and technical disciplines (STEM). This is indicated by the allocation of Research Council given by UKRI in the application assessment process. Currently the number of STEM fellows in academia outnumber the

number of fellows outside STEM (ie those in arts, humanities and social sciences or AHSS) by three to one.

We would therefore suggest segmenting the FLF cohorts into four, and to consider the most appropriate evaluation approaches and metrics for each. This is explored further in the following chapters.

**Figure 3-13: Fellowship breakdown**



Source: FLF management information, Rounds 1-4

### 3.5.2 Consideration of near misses

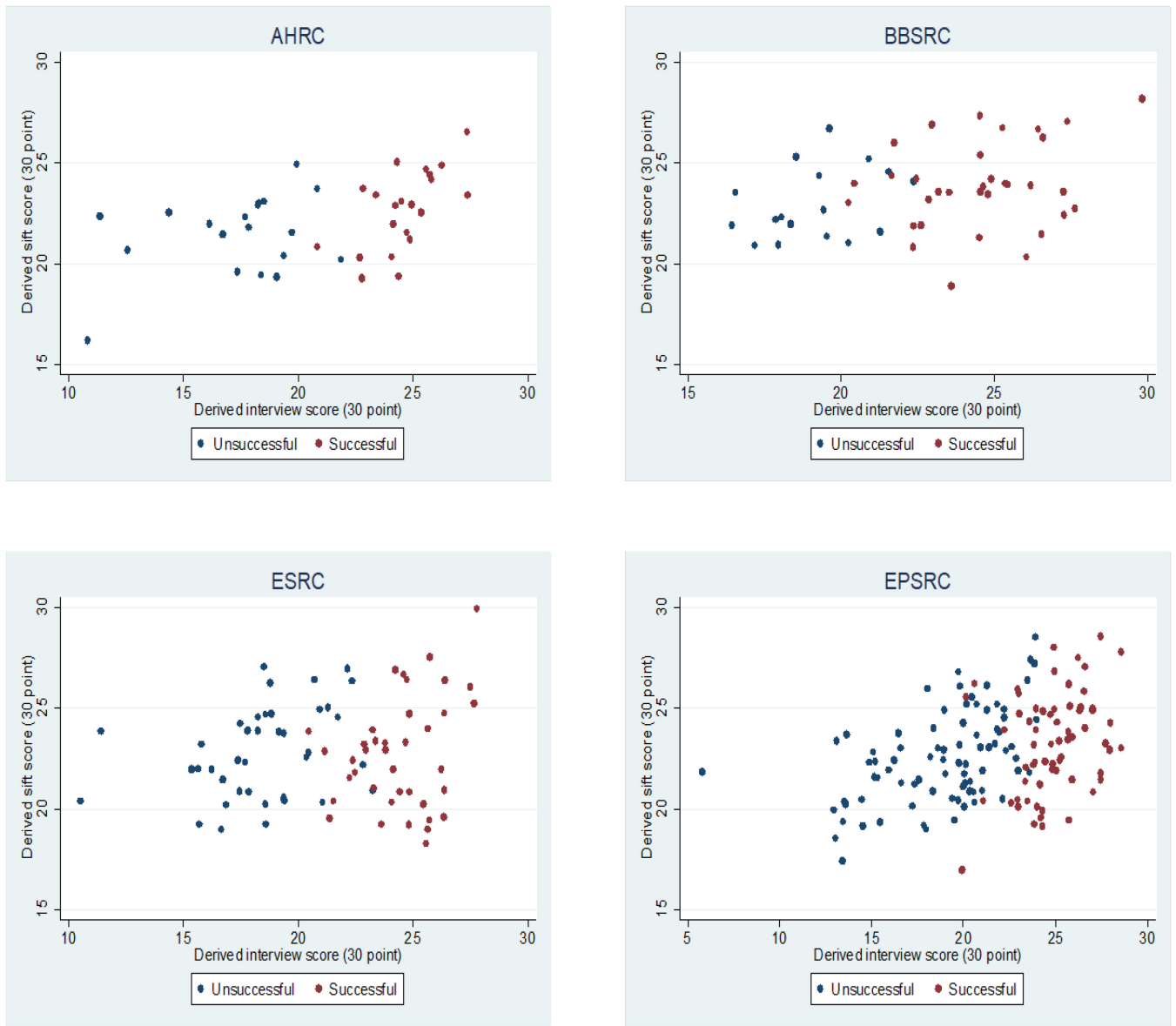
Consideration of evaluation methodologies includes the potential to use unsuccessful applicants as comparators for an impact evaluation of the FLF. A key issue is therefore to understand how closely matched unsuccessful applicants are to successful fellows in terms of motivation, and in terms of capability.

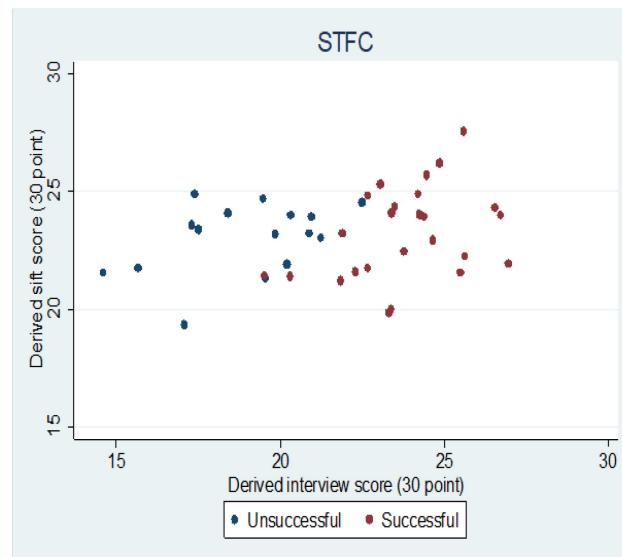
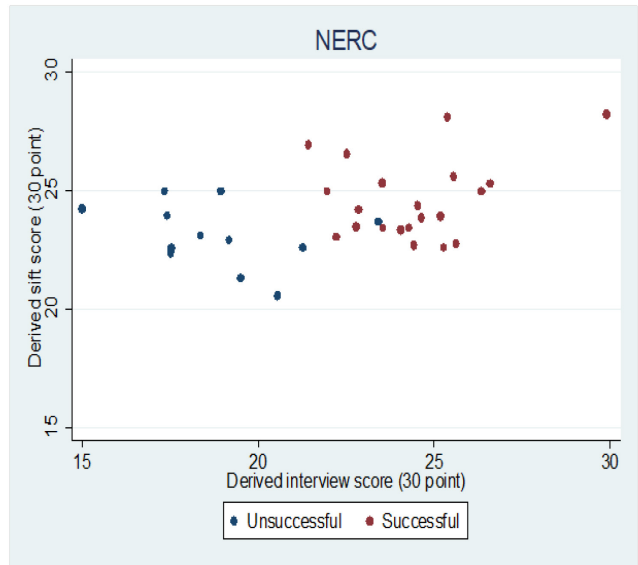
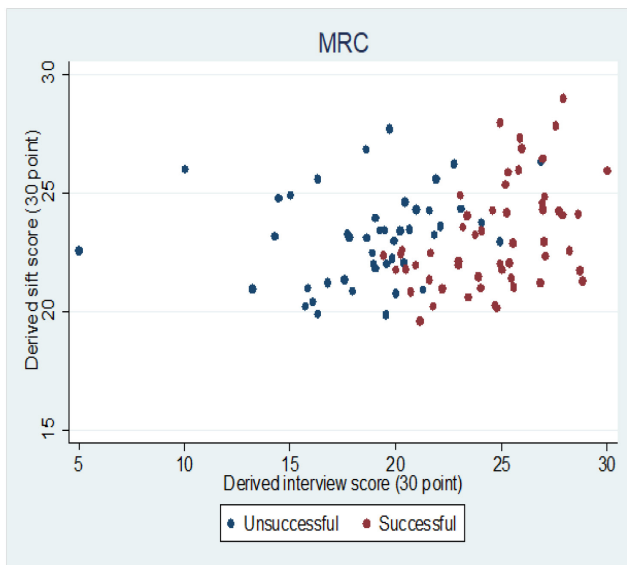
Motivation is difficult to determine in programme monitoring data (but can be explored in surveys and interviews). However, one can assume that the applicant pool as a whole contains highly motivated and intelligent individuals who, with or without dedicated fellowship funding and support, could go on to have successful and impactful research careers. The act of making an application, with the support of a sponsoring host research organisation, indicates a motivation and backing to develop a career in R&I. In terms of capability, this can be arguably determined using the peer-review scores produced during the FLF application process. Analysis of these scores indicates two additional implications for the evaluation approach:

1. The programme has become more competitive over successive rounds as it has attracted increasing numbers of applicants. For example, the data show that in the first two rounds there were no applicants who merited funding who did not receive it so applicants had a greater likelihood of success in early rounds compared to those who applied in later rounds.
2. Assuming the application process can accurately determine capability and potential, those who proceed to the latter stages of the process can be deemed to be more capable, and so those remaining and the scores achieved in these later stages should be compared. However, the distribution of scores varies by discipline. Plotting the sift scores against the interview scores across each lead council gives a sense of how close successful and unsuccessful applicants are (these measures are only available for those who reached interview stage).

This is explored further in the following chapters, and is helpful in explaining why PSM and RDD are the most suitable approaches for quasi-experimental impact evaluation of FLF (see Chapter 6).

**Figure 3-14: Plotting sift and interview score for successful and unsuccessful applicants by panel**





Source: FLF management information, Rounds 1-4



## 4 Evaluation considerations and challenges

This chapter summarises existing knowledge and information about potential evaluation methods as set out in the objectives for the Scoping and Feasibility Review. It considers the evaluation methods that could be used in a full evaluation of the FLF programme and discusses those that are likely to be most appropriate. It outlines the range of methods available and what to consider in selecting an appropriate approach; and examines the approaches that have been used to evaluate similar programmes and suggestions in evaluation frameworks developed around R&I to draw out lessons that can be learned for an evaluation of FLF. It also presents an overview of the key challenges that need to be taken into account when evaluating FLF and how they might be mitigated or accounted for.

### Key findings for an evaluation design for FLF:

- Few fellowship or similar programmes have been evaluated, and where this has happened: reports may not be publicly available, monitoring is largely limited to input data, and evaluation tends to be focused on the process in order to make improvements to programmes rather than on outcomes or impacts or is limited to fellows' surveys and case studies.
- It is relatively rare for impact evaluation aimed at establishing causal links to be attempted so the scope to learn from other evaluation work in this sphere is limited. There appears to be a move towards impact evaluation and the FLF evaluation as well as broader R&I ecosystem would benefit from sharing of evaluation methodologies and outputs including measures and datasets (with suitable permissions).
- Selecting an appropriate comparison group is critical to experimental or quasi-experimental evaluation, and where they have attempted this, evaluations of fellowship or similar programmes have largely drawn on unsuccessful applicants, where they can be identified/accessed, as they are deemed to be the closest match to awardees. This does require that baseline data is available (or can be found) for both awardees and applicants. This suggests an approach for the FLF evaluation.
- Other methods have been used to draw from a wider pool but this tends to have been driven by lack of information about unsuccessful candidates/applicants, is resource intensive and is acknowledged could lead to biased estimates of award effects due to motivational differences in the control and treatment groups.
- Evaluations of fellowships or similar programmes have tended rely on the collection and analysis of bibliometric information and to rely on just one or two sources of data which is cost effective but narrowly defines outputs and is likely to be biased to certain disciplines and career pathways. This may not be a problem for evaluations of programmes with a narrow focus or eligibility but will be a problem for FLF which has broad ambitions and remit.
- Evaluations also tend to rely on surveys of awardees and the information collected can supplement analysis of secondary data. However, as they are largely aimed at awardees they cannot provide a counterfactual, they rely on accurate and unbiased self-reporting, and there is attrition over time. In addition, many evaluations collect qualitative data which tends to focus

on experiences of the programme to support programme improvements rather than to identify and measure outcomes and impact.

- Evaluation across disciplines benefits from impact metrology (diversity of methods and multifaceted indicators), an approach which offers the potential to capture outcomes beyond those intended or expected, but this can be resource intensive; and interpreting results from metrics across diverse data subjects needs to consider ways to allow for fair and meaningful comparison.
- It is important to acknowledge that the effects and impacts of programmes such as FLF may not emerge until well after the funding period and so the length and timing of the evaluation period (not just the timing of data collection within the period) will influence the impacts uncovered.
- An evaluation of FLF faces a number of challenges including some specific to the broad scope and stretching ambitions for the programme and its fellows:
  - a) finding meaningful and robust ways to measure progress against high-level and wide-reaching change will require identification of intermediate and proxy measures that move beyond the narrow range of usual research intensity metrics and theory-based evaluation techniques to confirm pathways and linkages to impact;
  - b) identifying impacts resulting from FLF set against a landscape of other programmes will require understanding of the reach and impact of other schemes through sharing evaluation findings across UK R&I, and setting realistic expectations of what FLF can achieve which can be explored through interviews with strategic stakeholders;
  - c) understanding the outcomes and impacts achieved over and above what would have happened without FLF across the range of potential spheres of influence will require segmenting the Fellows into four sub-groups and adopting an evaluation method and metrics that are appropriate to the size and nature of each sub-group (eg, a counterfactual impact analysis using a comparison group for large segments, and matched case studies for the smallest segments can provide insights into additionality);
  - d) detecting spillover benefits and the amplifier or multiplier effect of Fellows on their teams, departments, host organisations and wider R&I culture will require theory-based evaluation approaches, identification of intermediate outcomes such as collaborations and networks and perceptions of influence which can be captured through surveys of Fellows and case studies;
  - e) understanding and accounting for wider external or contextual factors that can influence participation in and impacts of FLF will require creating a comparison group subject to the same environmental pressures as Fellows and working with high level stakeholders to chronicle the shifting landscape affecting UK R&I;
  - f) allowing time for impacts to emerge (and at different rates for different disciplines and MIDRI mixes) and expand out to be observed at levels beyond the individual Fellow will require an evaluation approach that extends beyond the funding period (up to 15 years from the start of the programme) and could involve annual sweeps of data collection and analysis to avoid imposing a timetable for outcomes to emerge which may be incorrect.
- Choosing an evaluation approach should be informed by the Theory of Change and understanding of the challenges faced and trade-offs involved; and stakeholders must be engaged to ensure a shared understanding of the limitations and what is feasible. This forms the basis of the scoping and feasibility study. Factors to consider include: the need for both accountability and learning so will need to understand causality; the stretching nature of the evaluation questions and likelihood of restricting the evaluation time period to the FLF funding period mean that not all the questions can be answered; and FLF is complex in terms of

inputs and activities (diverse fellows and flexible and adaptable support ) and in terms of desired outcomes and impacts, has diverse and interacting components and non-linear interactions, and complex systems such as FLF can be difficult to establish causal links. It will therefore need a mix of evaluation methods and techniques.

- The recommended approach to understand the impact of FLF would be to combine a full counterfactual quasi-experimental design (counterfactual impact evaluation) focused on fellows, with programme monitoring information (MI) analysis and theory-based evaluation to understand effects for hosts and through R&I ideas. Theory-based evaluation could mobilise career tracker surveys and matched-case studies. Additionally, a process evaluation is required to capture and document programme delivery.

## 4.1 Potential evaluation approaches

The evaluation questions set out for FLF by UKRI (see Chapter 1) indicate that three types of evaluation required:

- **Process evaluation** – relating to the appropriateness of the FLF programme. Process evaluation monitors programme activity to determine whether the programme is working and being delivered as intended, and what is working more or less well and why. It reviews activities and outputs, tracks programme progress, and picks up on early warnings of any issues that may occur; and tests and fine-tunes the theory of change. Understanding the ‘how’ of FLF is important to UKRI.
- **Impact evaluation** – relating to the effectiveness of FLF. Impact evaluation takes place during the runtime of a programme in certain periods or at the end of a programme and beyond. It helps assess if, and to what degree, the programme meets its goal and how well have the desired short-term (or long-term) changes been achieved.
- **Economic or value for money evaluation** – relating to the value for money represented by FLF. Value for money evaluation may inform whether the intervention was indeed a good use of resources given the benefits it produces and its costs: comparing costs and benefits. It can involve absolute and relative value for money - assessing whether benefits outweigh the costs, and whether cost effectiveness is greater than that achieved by other programmes. UKRI acknowledge that it will be difficult to derive a robust assessment of value for money based on an assessment of the value of impacts (and how these might be ‘monetarised’).

There are two main approaches to impact evaluation - experimental or theory-based approaches (although these are not mutually exclusive) – and the choice between theory-based methods and experimental methods depends foremost on whether it is possible to compare groups affected and not affected by the intervention (Magenta

Book, HM Treasury)<sup>10</sup>. The most robust forms of impact evaluation rely on experimental methods for which a credible, counterfactual is a necessity.

**Experimental and quasi-experimental methods** are used to identify and measure the net impact of an intervention. The main idea behind these methods is that impact can be measured by comparing the ‘counterfactual’ outcome that can be observed from a ‘control’ group with that of the intervention (or treatment) group; and this overcome the difficulties of attribution. However, for these methods to be implemented soundly, the treatment and control groups need to be identical (with the only difference being the intervention) or differ in ways that can be accounted for. This can be challenging to achieve. Experimental impact evaluations with a counterfactual mainly refer to *Randomised Control Trials* (RCTs) which tend to be regarded as the ‘gold standard’ in impact evaluation however it is not practical, desirable or ethical to implement trials in all contexts. RCTs establish a control group who, in other circumstances might receive the intervention, and tracks their outcomes relative to the treatment group. In contrast, *Quasi-Experimental Design* impact evaluations (or QED) methods draw a comparison group in data in some form. The group is well-matched in terms of observable characteristics, with techniques such as Propensity Score Matching (PSM), difference in differences (DiD), instrumental variables, timing of events, and regression discontinuity design (RDD) used to establish this. The method selected depends on whether an intervention can be randomised, the expected size of the effect and collected sample size, and the ability to establish a control group.

**Theory-based methods** are used to explore impacts by discerning the causal links that are likely to bring about change by an intervention. These approaches are less robust and can draw on qualitative and quantitative data. As such they are unable to give a numeric estimate of the effect size of the intervention. However, theory-based evaluations can address the extent to which change took place and why it occurred, taking into account the context of the intervention being administered. These approaches are particularly well-suited for the evaluation of complex interventions and complex environments – which, as indicated by the theory of change, describes the FLF programme. They are also suited to programmes where attribution is challenging (discussed below), as is the case with FLF. Some of the most highly regarded theory-based methods include: (critical) Realist Evaluations, Qualitative Comparative Analysis, Contribution Analysis, Process Tracing and Simulation Modelling; and others include Outcomes Harvesting, Most Significant Change, General Elimination Methodology, and Success Case Method. These can often be very resource-intensive and may not be suited to accountability driven evaluation. For a full exploration and explanation of evaluation methods scoped see Appendix C.

#### 4.1.1 Choosing an evaluation method

The evaluation literature highlights that selection of the analytical method should be informed by the Theory of Change and the uncertainties and assumptions that it identifies;

<sup>10</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/879438/MT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/MT_Magenta_Book.pdf)

and that stakeholders are involved in the selection process as well as the development of the Theory of Change to ensure that the outcomes they are interested in are captured, and there is a shared understanding of the limitations and what is feasible. This in essence has been the thrust of the scoping and feasibility study for FLF.

There is a wide variety of potential evaluation methods that can be used when designing an evaluation and so shortlisting methods for FLF requires considering specific requirements for feasibility, comparative advantage, and shortcomings. Aspects to consider are noted below; and indicate the importance of clearly defining the population, the place and the time period of the evaluation, as well as the purpose of the findings, and whether or not they can be generalised.

■ Who is the evaluation for and what is its purpose?

The evaluation of FLF has a number of key audiences and interested parties including: UKRI to understand whether the unique aspects of the programme deliver on its ambitious goals and to understand how it works; and BEIS as funders of the programme to understand whether FLF delivers value for money. The purpose of evaluation is about accountability but also learning so needs to understand causality.

■ Will it meet the time and budget constraints and availability of data?

At this stage the timescale and budget for an evaluation of FLF is unknown. The Innovate UK Evaluation Framework (Innovate UK, 2018) notes that the UK Space Agency have a guide for setting out considerations of budget, and that for large, complex programmes and/or uncertain outcomes a detailed externally commissioned evaluation with a budget of one to five per cent of the total programme is recommended. It is possible that the evaluation period may not extend beyond the programme period, which mean some outcomes cannot be fully captured. The evaluation is likely to draw on a range of data quantitative and qualitative sources including secondary and administrative data, programme data (including those captured during the FLF application process) and primary data captured directly from fellows and their hosts. Some data sources may have restricted access to those outside of UKRI which will need to be negotiated, and some may require permission for processing and arrangements for data sharing.

■ Will it answer (align to) the evaluation questions?

The evaluation questions are well developed but they are many and stretching, and it is likely that not all will be answered within a restricted time period. It is unlikely to be possible to identify binary answers to all impact questions but there are better prospects for addressing the causal pathways, where this is the case.

■ Is it appropriate for the specific characteristics of the intervention, its subjects and context?

Evaluation is often a complex problem involving complex systems with diverse and interacting components, non-linear interactions between components, and adaptation and learning by components in response to change. Complex systems can make evaluation challenging by: being sensitive to context, making it harder to establish

causal links, and with different components having a disproportionate effect on the outcome(s). The work undertaken to develop a theory of change for the FLF programme (see Chapter 2) indicates it has a degree of complexity, some in the intervention itself with its flexible and adaptable support, but mostly in the desired outcome and impacts. Adapting the level of robustness for different measures will mean broader insights can be generated.

Most evaluations use a mix of methods, both qualitative and quantitative techniques and data, for different purposes or to shed light on different aspects. The evaluation process itself is also likely to be iterative, with learning and reviewing occurring during the evaluation, and adapting to changes and details as they come. No method will be perfect, but some will have advantages over others in terms of appropriateness, timing, and cost.

Theory-based methods may therefore be suitable for an evaluation of FLF particularly: qualitative theory-based approaches that can help capture mechanisms of change and offer lessons and information for similar contexts/policies, and case-based approaches that help identify the combination of factors that are necessary for success in a certain context. However, this does not mean that: these theory-based approaches should be used in isolation, that quasi-experimental approaches cannot be used, or that the evaluation design should be fixed (as requirements can change and/or become clearer over time). The following chapters explore the feasibility and choice of approach(es) for each strand of the proposed evaluation design drawing on the theory of change presented in Chapter 2:

- counterfactual impact evaluation - Chapter 6;
- process evaluation and theory-based impact evaluation - Chapter 7; and
- economic evaluation - Chapter 8.

## 4.2 Learning from other evaluations

In considering an appropriate approach or mix of approaches for a full evaluation of FLF it may be possible to learn from other evaluations of similar programmes and initiatives. As discussed in Chapter 3, there are many fellowship and scholarship schemes aimed either at UK researchers and innovators or running in UK universities and businesses however few have been subject to evaluation.

### 4.2.1 Evaluation activity

The review found that few programmes (c.20, see Table 4-1) appear to have been evaluated or state that some form of monitoring and/or evaluation is planned or is underway. Where evaluation has or is planned this has tended to be programmes funded by UKRI or its constituent bodies or funded by health bodies. There were additionally a few examples of evaluations of non-UK based programmes. Key findings include:

- It is difficult to find published evaluations. In some cases, programmes have been monitored and reports are available; and, in others, programme guidance can explicitly state that data will be collected and held for monitoring and evaluation purposes but no full monitoring or evaluation reports appear to be publicly available.



- Monitoring has largely been limited to input level data such as numbers of awards, sums spent and, in some cases the characteristics of recipients and case studies of individual projects or recipients/fellows/ prize winners. This monitoring can demonstrate to trustees/key stakeholders how the funds have been spent but there is perhaps an unwritten expectation that the funder will indirectly benefit from this spending on research(ers) and innovation(ers) with no expressed causal link and no evidence of impact of the funds/programme etc. In most cases when exploring the nature of the programme (aims, eligibility, support etc) there is no mention of evaluation.
- In some cases, a formative process evaluation has been undertaken to understand the nature of the programme and particularly the support received, in order to make improvements for the future.
- In a small number of cases, there has been an attempt to undertake a summative evaluation of the impact of the programme but this has largely been limited to surveys of fellows and individual case studies rather than a full counterfactual impact study (using a control or comparison group)

Detailed findings from these evaluations are provided in Appendix D.

The lack of evaluation of fellowship programmes could be due to the complex nature of funding arrangements, particularly where there are joint or multiple funders; the scale of programmes, which can be quite small or very large; or the varied nature of the activities funded. However there appears to be a move towards impact evaluation of programmes, with some (particularly larger, UKRI and research council funded) programmes noting upcoming monitoring and evaluation. It would benefit the R&I ecosystem if evaluation outputs (and also evaluation methodologies and experiences including challenges faced) were shared more readily and openly; and, much more ambitiously, if a base set of common metrics developed and agreed (for example in tracker surveys). This would also fit with the work of Innovate UK which is trying to take a more holistic approach to evaluation (rather than programme by programme) and would allow the R&I system to better understand the impact of research funding.



**Figure 4-1: Overview of monitoring and evaluation methodologies used in similar programmes**

UK FELLOWSHIP PROGRAMMES (+ others)	Size	Survey	Qual	MI	Secondary analysis	Counterfactual	Economic vfm
RSE Enterprise Fellowships	244	✓	✓	×	×	×	×
Clinical Research Training Fellowship	231	✓	×	×	×	×	×
CLARHRC Fellowship	29	✓	✓	×	×	×	×
Dorothy Hodgkin/University Research Fellowship	1800	✓	×	×	×	×	×
Clinical Scientist Fellowship	21	✓	✓	×	×	×	×
Springboard	105	✓	✓	×	✓ bib	✓ random	✓
Sir Henry Wellcome Postdoctoral Fellowship (tracker)		✓	×	×	×	×	×
DigitalHealth.London Accelerator (SME support)	21	✓	×	✓	×	×	✓
ERC Starting Independent Research Grant (NotUK) ERC Advanced Investigator Grant (NotUK)	138	✓	✓	×	✓ bib	✓ Threshold + matching	×
Engineering for Development Research Fellowship	66	✓	✓	✓	×	×	×
Rising Stars (+Recapturing Talent, Research Fellowships, Chairs)	51	✓	✓	✓	✓ hesa	×	×
Global Challenges Research Fund (Project funding)	1410	×	✓	✓	✓ bib?	×	✓
International Research Fellowship Programme(USA)	1039	✓	×	✓	×	✓ PSM	×
ESRC Postdoctoral Fellowships	293	×	✓	✓	×	×	×
Independent Research Fellowships (NERC)	329	✓	✓	×	×	×	×
BBSRC Fellowships (David Phillips and Discovery Fellows)	~100	✓	✓	✓	✓ bib	×	×
EPSRC Postdoctoral Fellowships	~500	✓	×	✓	✓ bib	×	×
ICURe	~220	✓	✓	✓	✓ HE-BCI	✓ D-I-D	×
Erwin Schrodinger Fellowships (Austria)	2,271	✓	✓	×	✓ bib	✓	×
Emmy Noether Programme (Germany)	659	✓	✓	×	✓ bib	✓	×
Marie Skłodowska-Curie Actions (various)		✓	✓	✓	✓ bib	✓	✓

Source: IES, 2021

## 4.2.2 Lessons from the evaluations

### Developing a comparison group

An important part of an impact (experimental or quasi-experimental) evaluation is selecting the most appropriate comparison group. However, as indicated from the review of evaluations of early career awards (and similar), much of the evaluation work undertaken to date has not attempted a counterfactual impact study. Where this has been addressed the most common approach of choosing a comparison group has been to use the unsuccessful applicants, for example for the ICURe evaluation (Ipsos MORI, George Barrett & Tomas Ulrichsen, 2018), the review of the Springboard scheme (Freshney Consulting and Aleron, 2019), the evaluation of International Research Fellowships in the USA (Martinez, et al., 2016), and an evaluation of European Research Council programmes including Starting Grants (Nedeva, et al., 2012). These unsuccessful applicants are likely to fulfil the eligibility criteria for the grants and hence be similar to the awardees with respect to some main characteristics (for example discipline and career stage). This approach requires that some information on the unsuccessful applicants is available to the evaluators. However, in the

evaluation of the Marie Skłodowska-Curie Actions (European Commission, 2017) the names of the unsuccessful applicants were not available, although the names of their institutions were, and so those institutions were used to identify individuals who comprised the control group. By focusing on the unsuccessful applicants, the evaluators are likely to identify individuals who are similar to the successful applicants in terms of some observable characteristics and perhaps unobservable characteristics, like motivation. However, as application eligibility might allow for a wide pool of potential candidates to apply to the programme, not all unsuccessful candidates will be appropriate for joining the control group and, similarly, not all applications will be of equivalent quality.

Some evaluators recognised that not all unsuccessful applicants are similar enough to the awardees and set some criteria for choosing the most appropriate group. For example, in the evaluation of the Marie Skłodowska-Curie Actions (European Commission, 2017), only the institutes of applicants whose applications were above the quality threshold were eligible to be in the control group. However, as any evaluation is subject to sample size restrictions, imposing more exclusion criteria might lead to a better matched comparison group but with the cost of decreasing the sample size and the detection of the programme's effect. As both dimensions are important in the validity of the evaluation and need to be carefully considered.

The reviews of evaluations indicated that identifying a comparison group is possible even when there is no information available on unsuccessful candidates (meaning that unsuccessful applicants cannot be purposely selected as a control or comparison group). In these circumstances the evaluators sought to find individuals who were very similar to the awardees with respect to relevant characteristics. These could include aspects such as their discipline, career stage, job position, number of publications and citations during the pre-award period, type of institution, and even demographic characteristics, such as ethnicity and gender. In the evaluation of the Erwin Schrödinger Fellowships with Return Phase (Meyer & Bühner, 2014) a random control/comparison group was chosen comprising of individuals who matched the successful candidates with respect to their gender, scientific discipline, and publication age. Also in the evaluation of Marie Skłodowska-Curie Actions schemes (European Commission, 2017) a comparison group of researchers was created by identifying similarly successful established researchers in similar fields to their funded researchers. Both these evaluations used bibliometric analysis to develop the comparison groups. This approach could allow for a wide pool of potential candidates for the comparison group to be selected from but is likely to be time intensive. Another caveat of this approach is that the individuals of a randomly chosen comparison group might not be as motivated and career driven as the awardees, which could lead to biased estimates of the awards effects on the participants' future outcomes.

A small number of evaluations used a theory-based approach to impact evaluation hence did not identify a comparison group. This includes the evaluation of CLARHC East of England Fellowship Evaluation. Also, the evaluation of ESRC Postdoctoral Fellowships did not have a comparison group but sought to compare the outcomes of different groups of fellows within the programme.

## Data collection

In the evaluations reviewed – both those where a comparison or control group is compared to group of successful applicants/awardees and those focused on awardees/fellows – a common approach to retrieve information is the bibliometric approach. This entails the collection of publicly available information on characteristics such as number of publications, number of citations, career stage and employment position, other grants and awards, and possibly demographic characteristics such as gender and ethnicity. This approach is being used in ongoing evaluation of EPSRC Fellowships.

This is a cost-effective way to collect relatively robust information but places particular weight on dissemination outputs which may be narrowly defined (eg limited to published articles and papers in academic journals) and be biased towards certain disciplines and academic careers. This may not be a problem for evaluations of programmes with a narrow focus or eligibility but will be a problem for FLF which has broad ambitions and remit. Some of the main resources used are Scopus and ResearchFish (Meyer & Bühner, 2014; European Commission, 2017), but also others such as Web of Science (Hornbostel, et al., 2009), PubMed, and ORCID website (Freshney Consulting and Aleron, 2019). In some cases, the personal websites of the awardees and the individuals in the control group were also reviewed to capture relevant information but this could prove time consuming, especially if there are large numbers of individuals to track in this way and/or if repeated data collection is required (see Chapter 6 for a discussion of methods to establish a counterfactual group for FLF).

Another way to collect robust information on fellows/awardees is through tracker survey. A key learning point from the review is that most of the evaluations identified include some type of survey, directed either only to the programme participants or in some cases also to non-participants who are in the control group (see Chapter 7 for a further discussion of tracker surveys). These surveys could supplement analysis of secondary data or could provide the entirety of data analysed. Tracker surveys can be sent on a regular and rolling basis and allow individuals to be tracked over time to create a longitudinal dataset; or can be undertaken as a one-off or periodically across an entire cohort (who are at different stages in their careers and post programme support) to create a snapshot of a cohort and allow for analysis of career progression within the dataset. Tracker surveys however rely on the perceptions of and accurate reporting of awardees and are rarely externally validated. Tracker surveys were used as the primary source of data in evaluations of Royal Society Dorothy Hodgkin and University Research Fellowships (which compared aggregate results to the Principal Investigators & Research Leaders Survey), Clinical Research Training Fellowships, BBSRC David Phillips Fellows and Discovery Fellowships, AMS Clinical Scientist Fellowships, RAEng Engineering for Development Research Fellowships

In addition, most cases involve qualitative data collection through interviews with a sub-sample of the survey respondents to get better insights into the ways the programme facilitated the awardees career paths, problems that arose, and ways to improve the programme. For example, evaluation of the Royal Society of Edinburgh Enterprise Fellowships involved a survey of participants and case studies to explore how (well) fellows were supported, and how the programme could be enhanced; this was also the approach

taken in the evaluation of NERC Independent Research Fellows and Phase 1 of an evaluation of NCTL's High Potential School Leaders programme.

### 4.3 Learning from wider evaluation work

The review also identified other useful material and insights for FLF particularly that relating to the development of evaluation frameworks and approaches. These emphasise the importance of an underpinning logic model or theory of change (or multiple theories following different activities which identify mechanisms of change) and also highlight a number of other aspects/dimensions to consider:

- Need to move beyond solely process evaluation and emphasise the outcomes and impacts of funding (NAO, 2021) and to use a control group approach wherever possible. In the case of Innovate UK *'the high quality, unsuccessful group therefore becomes our most solid control group'* (Hodges, 2018). But to recognise that evaluation will never be perfect.
- Focus of data collection – should the approach be case study based as in the Research Excellence Framework (REF) approach to develop a strong narrative but which presents challenges for programme evaluation (Morgan and Jones, 2017), or metrics-based as in the recently developed Knowledge Exchange Framework (KEF) which draws on existing data largely from the HE-Business and Community Interaction survey to reduce burden on the sector/participants, or survey/reflection based as in the Vitae's Researcher Development Framework. The Innovate UK Evaluation Framework acknowledges that evaluation approaches generally need a mix.
- Need for impact 'metrology' (Jones et al, 2017) – a diverse array of methods and multifaceted indicators is needed to work across disciplines, a 'one-size fits all'/standardised approach or set of measures may not be appropriate or desirable.
- Normalising results – in implementing the KEF it was considered important to create clusters of HEIs (data subjects) to allow for meaningful and fair comparison or benchmarking. In this approach, data subjects are clustered into similar groups taking into account size, discipline base, and size and source of research income.
- Potential to look beyond intended and expected outcomes - different types of impact can emerge from programmes beyond the expected. Nedeva et al (2012) differentiates 4 types of impact: straight-runs (intended and expected outcomes) when most impact studies focus; plus long-shots (intended but not expected), collateral (unintended but expected), and accidental (neither intended nor expected). The authors note how impact studies tend to focus on straight runs.
- Timelines – the effects on researchers, research content and research careers can start from grant application (before programme or funding) but the full impact is not likely to be apparent until well after the funding period, and so the length and timing of the evaluation period is likely to influence the impacts uncovered (Nedeva et al., 2012). For example, scoping work for an evaluation of the Global Challenges Research Fund (Barr et al., 2018) proposes an evaluation that runs for at least 10 years in recognition that

pathways to impact are long, taking 10-15 years for impacts to emerge, and that over time impact appears at levels beyond individuals/individual awards.

## 4.4 Evaluation challenges

The literature review, review of other evaluations and stakeholder engagement highlighted that an evaluation of FLF faces a number of challenges. Many of these are shared by other evaluations of R&I funded initiatives and programmes including other early career researcher fellowships, but some are specific to FLF's diversity and broad scope in terms of participation and objectives which extend beyond usual research intensity measures.

These general and more specific challenges need to be recognised and taken into consideration in designing an evaluation. The challenges, implications for an evaluation design for FLF (ie why they matter) and how these could be addressed are summarised below along with approaches that could be used address them.

### Determining whether FLF has met its objectives

1. **Challenge:** As set out in the theory of change (Chapter 2), the programme has a range of ambitions for high level and wide-reaching change in R&I: culture, ways of working, make-up of talent-base, and leadership; and for impacts to ripple outwards from the individual to influence their research groups, research organisations, disciplines or sectors, and national or international contexts. Although it is not expected that all Fellows will meet all these ambitions, this creates challenges to find meaningful and robust ways to measure progress against these broad aims and levels of impact. The programme also has key objectives around: developing, attracting, retaining and sustaining R&I talent in the UK; fostering new career paths and mobility; and providing sustained and flexible funding for novel and ambitious programmes. These objectives are focused on the individual fellow so determining whether the objectives have been met will also need to focus strongly on the level of fellow.
  - a. Implication for the evaluation: There is a need to identify relevant outcome measures and proxy measures where outcomes and impacts are difficult to measure or cannot be directly observed. The theory of change for FLF sets out pathways and mechanisms for change driven by assumptions – identifying outputs, outcomes and intermediate impacts which could lead to wider system and societal level change. For example, it is assumed that a good R&I leader can be a catalyst for cultural change (wider impact) by embedding good leadership values and practices, and facilitating opportunities for those around them to develop and flourish (intermediate impacts); and it is assumed that diversity among FLF participants (outcomes) can lead to diversity in their research groups and organisations (intermediate impacts) which in turn leads to increased learning and innovation (wider impact).
  - **Solution:** Theory-based evaluation approaches should seek to trace and confirm these pathways and linkages to impact, and capture data on perceived progress towards outcomes including for those that are less tangible; this will allow for hard measures to focus on outcomes and intermediate impacts (Chapter 7).

- b. Implication for the evaluation: There is a tendency in evaluations of research programmes to stick to a narrow range of ‘usual’ research-intensity metrics e.g. published articles and citations. However, the ambitions and scope of FLF are much broader, the UKRI is a signatory of the Declaration on Research Assessment (DORA) which explicitly seeks to improve the way research outputs are evaluated. Stakeholders are keen to ensure an evaluation looks beyond narrow bibliographic information measures to capture the wider effects of FLF.
- **Solution:** It will be important to capture a wider range of research outputs to reflect the FLF career pathways and contributions. Recommendations include: information on collaborations; tracking contributions of individual fellows; ways to identify multi- and interdisciplinary research and innovation (MIDRI) including working across sectors and disciplines; and measures for good leadership (drawing on Vitae’s Researcher Development Framework and its Leadership Lens and related tools). Analysing a wide range of measures will help to ensure these are inclusive and work for individuals from a wide range of backgrounds and settings (Chapter 5).
- c. Implication for the evaluation: The FLF is also expected to deliver value for money. This is challenging to evaluate, and an evaluation design will need to establish expectations around overall level of financial returns that would indicate value for money, the acceptable level of risk associated with challenging/innovative research, timeframes of expected returns, whether non-financial impacts can be monetarised or considered value for money (eg policy change), and whether FLF can be compared with other programmes or funds in a meaningful way (see Chapter 8).
- **Solution:** Economic evaluations can take a number of forms, which range from intervention-level to multi-stranded programme level, and it will be necessary to mobilise several strands to fully understand the returns to FLF, given the different evaluation approaches that are viable to measure impact and different pathways to impact. It may prove more feasible to measure absolute value for money, rather than relative value for money. As economic evaluations are undertaken in the end-stage of multi-phase complex evaluations, we recommend a mid-term review and strategic assessment of outcomes to inform ways forward for FLF.

### Attribution of impact to FLF

2. **Challenge:** A core challenge is to be able to appropriately identify impacts resulting from the Fellowship (what is plausible to attribute to individual fellows and the cohort of fellows; and also disentangling the impact of the Fellow from their research), and this is exacerbated by the breadth and scope of the FLF and the wider context of funding programmes.
- a. Implication for the evaluation: Attribution is challenged by other programmes having similar aims and participants – such as aiming to facilitate collaboration between academia and industry, encourage multi- or interdisciplinary-working, facilitate international mobility, and/or develop future R&I leaders. There are also funding programmes with similar, substantial levels of support (time and resource) (see Chapter 3 and Appendix B). The FLF is therefore not the only programme seeking to make significant impacts in R&I and fellows are likely to be working in fields and

sectors where other programmes are seeking to have an influence. Thus, identifying the contribution of any single programme with observed performance improvements is difficult (Innovate UK, 2018). These wider opportunities will also complicate estimations of additionality (see below).

- **Solution:** The potential influence of other schemes and programmes has to be considered in the evaluation design. While, it might be judged that drawing a counterfactual for impact evaluation from other, similar programmes might help identify the unique contribution of FLF, there would be downsides in respect of data collection and unobservable characteristics, hence, to understand net impact, drawing a control group from within the programme is optimal. These disadvantages are minimised within theory-based evaluation, where benchmarking FLF fellows and hosts against other programmes (through surveys - assuming common metrics are used, and/or using secondary data analysis e.g. of Researchfish) will help to contextualise progress observed on causal pathways.
- b. Implication for the evaluation: Attribution is challenged by the relative scale of FLF, and the complexity of the R&I environment. Over three years, 550 Fellowships will be awarded but the number of early career researchers and of researchers and innovators across academia, industry and government and the third sector is considerably larger. To what extent can this small number of individuals be expected to bring about change in the wider R&I ecosystem, and what can be traced back to them? Also, whilst it is relatively straightforward to identify intermediate impacts of the Fellowship in the first few years (whilst receiving support) it will become harder as Fellows progress through their careers, to discern what can be realistically be attributed to the impact of the FLF programme. Finally, it may be difficult to identify and attribute the impact of individual fellows in different settings. For example, for Fellows based in industry it may be challenging to detect impact at the host level when Fellows are based in large corporate organisations with substantial workforces in a number of countries.
- **Solution:** The theory of change identifies the causal pathways to intermediate impacts that underpin final impacts. These should take less time to emerge than the final, system level impacts. Nonetheless, it will be necessary for the evaluation timeline to allow for differential timings of outcomes for different disciplines and MIDRI mixes. The different timelines to impact within different disciplines and MIDRI mixes can start to be described, if not fully elaborated, in the period of the FLF evaluation through tracing through the causal pathways, and regularly seeking to observe the selected outcome metrics. Using benchmarking in the theory-based evaluation will add contextual information based on the outcomes achieved by other similar programmes. Strategic assessment by senior stakeholders, based on evaluation evidence, will be important to help to identify the contribution of FLF to the landscape for R&I.

### Additionality of FLF

3. **Challenge:** A key ambition for the evaluation of FLF is to give an indication of the outcomes and impacts achieved over and above what would have happened in the absence of the programme ie additionality; or indeed for outcomes to have been



achieved over a shorter time period and/or at a higher level of quality (ie partial additionality). Detecting partial additionality may be a more reasonable ambition for an evaluation of FLF given the likelihood that those who can or do apply but are not awarded an FLF Fellowship will successfully apply for one of the many other fellowships or similar support packages. Additionality in evaluation is generally tackled with counterfactual impact analysis using a comparison group, in this case, individuals not awarded a Fellowship, and to compare outcomes of Fellows and non-Fellows.

a. Implication for the evaluation: The diversity of Fellows and applicants (the characteristics and experience of Fellows and their focus and sites for R&I); and the broader landscape of funding and support adds significant complexity to establishing baselines and identifying counterfactuals. This particularly complicates any identification of a counterfactual group *external* to the programme that could embody all of the characteristics of the FLF Fellows including their motivations, within a similar timescale, opportunities for obtaining funding and research journey.

- **Solution:** Building comparison groups from within the programme is likely to be the best way forward ie building a control group from applicants who are not awarded an FLF fellowship. It is likely to be important to divide the population of applicants and Fellows into segments relating to sector and then discipline or host size (so far as the data allows) to ensure the best matches are achieved between the applicant and fellow group e.g. STEM academic, AHSS academic, SME business, and large business. This will in likelihood lead to challenges with some groups too small for robust, counterfactual quantitative approaches and analysis, where instead a theory-based mixed method is needed, allowing for outcomes to be traced relative to the causal pathway to understand the likelihood that impacts will emerge. Secondly, there is opportunity to mobilise less robustly-matched control groups through using a peer case-study approach drawing on judgements within academic databases and then collecting detailed information on matched peers relative to fellows. This would provide further insights into additionality particularly where sample sizes mean that counterfactual impact analysis will struggle (see Chapter 7).

If a comparison group was drawn from unsuccessful applicants (ie a group internal to the programme) this overcomes many of these challenges, and follows the work of other successful programme evaluations. This does require relevant data permissions and consents to be in place to process (and if appropriate) collect personal data; notably to process application data to provide a baseline and to enable close matching.(see Chapter 6).

- b. Implication for the evaluation: Additionality can be affected by wider aspects including deadweight, displacement and substitution. Deadweight occurs where the desired outcomes would have happened anyway without FLF, and there is an inherent assumption outlined in the theory of change that unsuccessful applicants to FLF can develop their idea without the FLF funding package but will not achieve as much. Displacement and substitution occur where the positive impacts of FLF are offset by negative outcomes elsewhere such as the scaling back of other fellowship schemes reducing wider opportunities or the requirement for Fellows to be offered open-ended contracts reducing permanent research vacancies in their hosts.

- **Solution:** Given the complexities of FLF, multiple solutions are required, because the same approaches cannot be mobilised for all groups (e.g. segments), R&I (discipline mixes) and host (academic or business). Using a comparison group from within the programme will enable the impacts of FLF to emerge relative to not receiving FLF (but also possibly receiving other funding). The theory-based research offers the potential to explore wider aspects affecting additionality to understand their potential for influence outcomes (see Chapter 7).

### Spillover and diffusion

4. **Challenge:** Additionality can also be affected by leakage, where the benefits of FLF leak to those outside of the programme, benefits that might be hard to predict or indeed to measure. However, this wider influence is an aim of FLF and can be deemed positive 'spillover' or diffusion of positive behaviours so is not a challenge per se. For example, as outlined in the theory of change, it is anticipated that Fellows working in teams will model good leadership and facilitate an inclusive and collaborative culture which will impact upon their research teams, who will then also model this behaviour leading to wider influence. Also it is anticipated that Fellows will engage in collaboration and create networks and linkages which will erode silos and increase porosity, and will enable their hosts to continue to work in these new ways. These anticipated spillovers are noted in the theory of change as the 'multiplier effect'. What is challenging however is to capture these spillovers/diffusion.
  - a. Implication for the evaluation: Is it possible to monitor impacts on others such as research team members and hosts, and to take account of the differing opportunities for influence/spillover depending on Fellows operating environments and research focus?
  - **Solution:** The potential for these multiplier effects should be tested in the theory-based evaluation, and intermediate outcomes which lead to multipliers can be captured such as number and nature of collaborations, and persistence of partnerships through secondary data and surveys of Fellows and hosts. Additionally, qualitative case studies (nested, involving PI, the R&I idea and host, and potentially the research team) can create in-depth pictures of how individual fellows are influencing their environments and creating multiplier effects (see Chapter 5 for a discussion of metrics and Chapter 7 for recommendations for theory-based evaluation).

### Externalities

5. **Challenge:** A further challenge for an evaluation of FLF is to understand and account for wider external and/or contextual factors that can not only influence outcomes and impact, but also inputs (eg the likelihood of individuals applying and succeeding in being awarded a Fellowship). This can be referred to as endogeneity.
  - a. Implication for the evaluation: If these external factors are not controlled for, evaluation findings may over or understate the impact of FLF. Contextual factors include policy shift, new government, new leadership in UKRI, the UK leaving the EU

and most recently the COVID-19 pandemic. The challenge is being able to observe and measure these factors.

- **Solution:** A counterfactual impact evaluation involving a comparison group drawn from ‘within’ the programme will help to mitigate the need for measuring and including contextual factors in analysis as both Fellows and the comparison group will have experienced these events at the same time. Understanding the context and how it changes over the life of the programme, and situating FLF within this shifting landscape remains important, particularly to understand whether wider factors will have differential impacts on groups within the programme. This is a key thrust of the process and theory-based evaluations approach discussed in Chapter 7, involving operational assessments by those involved in delivery and strategic assessment from high level stakeholders.

## Diversity

6. **Challenge:** FLF has a wide scope and eligibility which sets it apart from many other fellowship programmes and this also creates challenges for evaluation.
  - a. Implication for the evaluation: There are particular challenges in identifying measures and information sources that would be applicable to all participants and applicants; in allowing for different career trajectories and differing timelines for impacts to emerge, and in establishing comparable groups for a counterfactual evaluation (as discussed above). Analysis of the programme monitoring data (presented in Chapter 3) indicates there are clear trends in applications and in award rates with greater volume and success rates among those applying from academia and with STEM focused research experience and proposals.
  - **Solution:** It will be important for the counterfactual impact evaluation to draw from a range of sources and to track outcomes over an extended period to allow comparison across the cohort, and for the theory-based impact evaluation to investigate whether and why the FLF is more effective for particular groups of individuals.

## Timing

7. **Challenge:** The vision for FLF Fellowships is ambitious, encompassing not only significant impacts for the career and outputs for individual Fellows but also expanding out to research/innovation teams, institutions, and across sectors. A long time may be required to effect change or observe the impact of changes at a higher level around culture, career pathways and workforce, as well as to embed these changes. In addition, the innovative nature of the research undertaken by Fellows means it may take longer to reach a stage where findings can be disseminated or patents registered. A further issue is that Fellows are likely to be highly mobile (geographically and sectorally) and so tracking them over time in data sources or with primary data collection such as surveys may be difficult.
  - a. Implication for the evaluation: This presents two key considerations for an evaluation: when is it meaningful or possible to measure which impacts, and how to monitor or collect relevant information over the longer timeframe of an individual’s career. Taking

a short-term focus or focusing solely on research outputs may miss key activity, however taking a broader and long-term approach to evaluation will require significant investment. Again, this presents a trade-off that commissioners will need to face when making decisions about the nature and extent of evaluation they are prepared to fund.

- **Solution:** A range of outcome measures and intermediate impacts are suggested for the evaluation (outcome/impact ‘metrology’), along with a suggested timeline of when these are likely to emerge short, medium longer-term (see Chapter 5). However, it will be important for the evaluation to track individuals and capture outcomes and impacts beyond the initial funding period which will require a longer-term commitment of resources. This longer-term view will also allow for developments in data collection around R&I activities and outputs to be factored into the evaluation, and enable measurement and tracking of information that is not currently possible to collect.

It may also be beneficial to track the full range of outcomes identified on an annual basis (which in the case of Fellow surveys will require engagement activity but should lead to lower attrition). This will allow for subgroups within the FLF cohort to ‘catch up’, provide additional evidence on the timing for outcomes to emerge for example by discipline and/or location of host, and avoid imposing assumptions on data collection which means outcomes are missed or not captured as they emerge. There is a trade-off to be considered here between costs, resource intensity and potential burden on respondents, and the quality and completeness of data collection. Making use of secondary data sources could minimise burdens of tracking and some costs.

However, another challenge for timing is the time-lag inherent in data sources, in terms of the frequency of updates. This can be mitigated to some extent through analysis of secondary data on an annual basis, and establishing clear notes about the periods covered by the data utilised.

## Data collection

8. **Challenge:** Evaluation activity of other fellowship or equivalent programmes has tended to involve surveys of awardees and/or analysis of external secondary data sources to collect (and less frequently to compare) outcomes and impact information. Data collection and analysis can involve challenges and some degree of trade-off.

- Implication for the evaluation: Data collection and analysis will incur financial costs: costs to administer surveys if used (although this can be greatly reduced through web-based survey administration) or costs to access data from third-party organisations. Also, care must be taken to consider the quality of the data: i.e. reliability and relevance of secondary data sources including if they are based on self-report, are biased towards certain discipline(s), and issues of disclosure (particularly involving intellectual property); and to consider the quality and bias in survey responses, as well as the burden on respondents and the sector.
- **Solution:** Drawing on a range of data sources rather than relying on one key source, as well as using a combination of primary data collection with secondary data analysis of administrative data will help to mitigate against concerns about quality and scope as well as burden (Chapter 5 considers and recommends the sources to use). Additionally, it may be beneficial for funders to share data (within the limits of

GDPR and with suitable data sharing agreements) on the programme participants and outcomes to increase the reach of monitoring and data that can be analysed, and to enable comparison of the performance of different programmes. This will need to be brokered by UKRI.

## 5 Outcome metrics

This chapter draws from the Theory of Change which identifies layers of outputs, intermediate and longer-term outcomes and impacts. It reviews and suggests a series of outcome metrics which could be used to evaluate the Future Leaders Fellowship fund. It also identifies potential data sources that can be used to observe the metrics and considers timing of when outcomes could be observed.

A complex programme such as FLF requires a toolbox of measures to fully capture the range of outcomes expected to emerge. FLF needs to capture outcomes for fellows and hosts that also relate to R&I and EDIP. The scoping and feasibility study has identified numerous measures to select from, with good prospects for capturing the effects of FLF, but further scoping will be required once information is available on the duration of evaluation and the timescale over which outcomes will be monitored as this will inform the utility of those noted below.

### Key findings for an evaluation design for FLF:

- The potential measures fall into eight broad categories, related to the theory of change. These ensure a wide and comprehensive set of metrics are considered thus moving beyond the 'usual' bibliographic indicators (supporting the principles of DORA) and ensuring applicability to the full spectrum of disciplines covered by FLF. The categories are:
  - a) Dissemination of research
  - b) Networking, collaboration and influence
  - c) Research funding
  - d) Business performance
  - e) Spin-offs and entrepreneurship
  - f) Patent and licensing activity
  - g) Leadership
  - h) Wider contributions.
- In each category a number of measures are suggested to form a long list of potential metrics. This will need to be pared down by the evaluation team as part of the final evaluation plan, and so will require further investigation (including test extraction) and stakeholder agreement (depending on budget) to provide a definite list of measures.
- Not all the suggested measures will be applicable to the entire FLF cohort. This is in line with the recommendation to segment the Fellows. It means a standard core set of measures could be used across the entirety of the population, with additional measures/metrics tailored to specific segments. Many of the suggested metrics can be observed at the level of the individual as the unit of assessment and can support a counterfactual impact evaluation. Some can be observed at multiple levels and others can only be observed at aggregated levels (often an artefact of data publication strategy) – for the department or cost centre (which gets closer to the research team activity), whole host organisation, groups of hosts, or sector. These aggregated metrics are useful for theory-based evaluation approaches.

- The data for many of the suggested measures can be obtained from administrative sources (and in some cases triangulated from multiple sources if the evaluation budget allows). These administrative data sources fall into three groups: individual level demographics, information relating to research, and information relating to the organisation (or host). However, information for some outcome and impact measures will need to be gathered through programme management information (MI) and primary data collection with Fellows and potentially their hosts via surveys and/or interviews, particularly information relating to career progression and trajectories, attributes and activities exemplifying good leadership, and feedback on the FLF programme. These data will fit better with theory-based evaluation approaches which do not rely on a comparison group.
- The outcomes captured by the suggested measures/metrics are likely to emerge (and have relevance) at different time periods driven by the nature of research and innovation (eg 'J' curve effect), host strategies, differences across disciplines, and artefacts in data collection and publication of data sources. An evaluation period of a minimum of ten years is recommended (15 years would be desirable) to allow for outcomes to emerge and be measured and to avoid underestimating the impact of the FLF programme. This can be divided into: early impact assessed within three years of receipt of the Fellowship; medium-term impact, between three and years; and longer-term impact, six or more years after the award. This will allow the process by which core funding shapes career development of individual Fellows, the context they operate in and also wider organisational impact and external reach/influence to be mapped out.

## 5.1 Categorising the measures

Measuring the impact of an intervention requires considerable thought about:

- a. the relevance of each potential outcome measure,
- b. what the precise nature of the measurement will be (how do we quantify it if we can),
- c. the unit of assessment or level at which outcomes are likely to be observed – is it the recipient of the intervention, their working group or environment or their organisation - and,
- d. the appropriate time horizon over which each potential outcome measure might realistically manifest itself.

Following the themes outlined in the Theory of Change the measures can be regarded to fall under the following broad categories. These are shown in Table 5.1, which also details the interaction with the TOC pathways and outcomes.

**Table 5-1: Themes and inter-relationship with the theory of change (TOC)**

Theme	Relationship with TOC	Outcomes measured
Dissemination of research	intermediate impact in the TOC on the fellow pathway	influential international research leaders
Networking, collaboration and influence	intermediate impact measure concerned with the fellow pathway EDIP mechanism	inclusive highly skilled R&I leaders



	host pathway	increased diversity and porosity more, new collaboration
Research funding	intermediate outcomes for fellow pathway the EDIP mechanism host pathway	influential international research leadership ongoing collaboration and porosity sustained role in R&I
Business performance	final outcome intermediate outcomes for host organisations	high performing economy sustained position in R&I
Spin-offs and entrepreneurship	final outcome intermediate outcomes on fellow pathway intermediate outcomes on R&I pathway host pathway	high performing economy research leadership leading edge R&I sustained role in R&I
Patent and licensing activity	Final impact R&I pathway intermediate and final impacts	high performing economy Leading edge R&I, ground-breaking R&I
Leadership	intermediate outcomes for fellows EDIP mechanism	influential international research leaders overlay between leadership and EDIP
Wider contributions	Final impact	Benefits to society high performing R&I ecosystem

Source: IES 2021

Each of these broad areas is addressed in turn and each contains a summary table of relevant metrics relating to the environment the recipient operates within (ie. the segments identified for the evaluation). At the broad level this is academia and business, but we consider further disaggregation. This reflects the fact that within academia there are very profound differences across broad disciplines (for example between STEM and other disciplines of Arts, Humanities and Social Sciences AHSS). Also within businesses an individual working within an SME faces a very different environment to an employee within a large corporation. The measures have been selected based on discussions in the two metrics workshops held with UKRI, a review of existing evaluation, and the scoping of potential data sources outside those commonly usually used for evaluation by UKRI and associated research councils.

Each summary table is organised in the same way:

- Suitability: firstly, each metric is reviewed to assess its suitability for evaluating impacts for fellows employed in academia, business or either type of institution. Where a metric is suitable for measuring the impact of the fellowship in a business setting it is reviewed to see if it is appropriate for those in small or medium enterprises (SMEs) or large firms or both. If a metric is suitable for assessing the impact of fellows working in academia, it is reviewed to see if it is suitable for those within a science discipline (STEM), or those



outside of science (AHSS) or both. This follows the segmentation of fellowships suggested in Chapter 3.

- Unit of assessment: as impacts are expected at the level of the individual, department or cost-centre and the host, this is considered for each metric.
- Data source: the next column refers to suitable data sources which can be used to assess the impact of FLF, this is following a review of data sources.
- Evaluation: the final column suggests the possible evaluation method or methods which can be used to detect impact for each metric. There are four key proposed methods: counterfactual impact evaluation, career trackers survey, theory-based impact evaluation (eg matched-case method) and programme MI data analysis.

## 5.2 About the administrative data sources

The scoping and feasibility review sought to identify possible secondary data sources that could be used to estimate the impact of the FLF programme on individual fellows, host organisations (businesses or academia) and wider society, and could potentially be available to an evaluation team. As well as considering where there might be gaps in secondary data and programme management data that would need to be filled with primary data collection (eg through surveys as part of the evaluation) The key administrative/secondary data sources identified are listed in the table below (Table 5.2) and categorised as to whether they are useful to provide demographic details of individuals, information about host organisations, or information about the research (and research outputs) (see Appendix E for a fuller description of the potential data sources and some of the advantages and limitations, although this is not a full critique).

It is worth noting here four key reflections of stakeholders on administrative data sources for evaluation which highlight how a one-size-fits-all approach is not suited to the complexity of the programme aims and structure:

- UKRI is committed to supporting the principles set out by the San Francisco Declaration on Research Assessment (DORA) which is against the use of journal-based metrics such as journal impact factors as a surrogate measure of the quality of individual researchers and their contributions (see Appendix A). Thus, bibliographic data sources such as Web of Science/Data Citation Index should be used with care - when it comes to published output the content of a paper is considered more critical than publication metrics, and bibliographic information should be supplemented with other data and measures (of success, influence or impact). Some potential data sources such as H-index and i10 index which are citation-based metrics that rely on journal impact and impose a hierarchy of journals should be avoided.
- There will be large differences by discipline in terms of the outcomes and perceived benefits produced and the timelines involved and so the evaluation will need to be inclusive and mindful that 'value is relative' and to understand the potential biases or coverage of certain data sources.

- An evaluation of FLF will need to consider at what level information is meaningful and most useful: macro-level, system-wide data versus micro-level data which includes information about individuals and employers. It was generally felt that data pertaining to individuals are the most valuable for an evaluation. Data relating to employers could be useful but there were concerns about the ability to detect the impact of FLF/evidence of change at this level (although this may be more visible at a research-intensive HE provider).
- Administrative data sources will not provide all the data required for an evaluation and that primary data collection through surveys, interviews or case studies of individuals and potentially their hosts will be needed (see Chapter 7).
  - Primary data collection would be most suitable for collecting information on career progression and trajectories. It would also be needed to collect other outcomes not covered by secondary sources such as: numbers of students or staff supervised, collaborations, intellectual property that was not patented, training or professional developed accessed, success rates of applying to grants after the Fellowship, promotions, and project outcomes in comparison to stated objectives.
  - Interviews with a subset of Fellows and could also be an opportunity to collect feedback regarding the FLF programme, such as the selection process or management of their grant.
  - Surveys of Hosts (academic departments/groups, institutes and businesses) could cover much of the same information, as well as staff retention, research group establishment, spill-over effects to hosts, and possibly validation of achievements. Case studies of hosts may be more appropriate for exploring change (and impact) from FLF particularly outside of research-intensive HE providers. These could tackle questions about how businesses get things done, how people are behaving differently as a result of FLF

**Table 5-2: Potential administrative data sources for an evaluation of FLF**

Type	Data source
Individual	<b>HESA Staff record</b> (Higher Education Statistical Agency)
Individual	<b>Longitudinal Employment Outcomes</b> (LEO)
Organisation	<b>Inter-Departmental Business Register</b> (IDBR)
Organisation	<b>Business Population Estimates</b> (BPE) (draws on IDBR)
Organisation	<b>Business Structures Database</b> (BSD, from IDBR)
Organisation	<b>Financial Analysis Made Easy</b> (FAME)
Organisation	<b>HE Business and Community Interaction (HE-BCI) survey</b>
Organisation	<b>Longitudinal Small Business Survey</b> (LSBS)
Organisation	<b>SME Finance Monitor</b>
Organisation	<b>HESA income data</b>
Organisation	<b>Community Innovation Survey</b> (CIS)
Organisation	<b>Beauhurst</b>

---

Research	<b>Cipher.ai</b>
Research	<b>Research Excellence Framework (REF) submissions data</b>
Research	<b>Dimensions database</b>
Research	<b>Researchfish</b>
Research	<b>ResearchGate network</b>
Research	<b>Scopus</b>
Research	<b>Web of Science</b>
Research	Individual HE provider <b>Online Research Archives</b>
Research	<b>Google Scholar</b>
Research	<b>RePec (Research Publications Economics)</b>
Research	<b>Altmetric.com</b>

---

### 5.3 Dissemination metrics

Table 5.3 introduces research-based metrics which are largely focused on the traditional academic peer reviewed research outputs. These types of measures are common and notably form a large part of the Research Excellence Framework (REF) in the UK (60 per cent in the 2015-2021 assessment) and similar exercises in an expanding number of countries. The UK's Research Excellence Framework (REF) is a system that is designed to evaluate the quality of the research produced by higher education institutes (HEIs). It focuses on three areas: quality of research outputs; impact of the research beyond academia; and the research environment. This initial assessment on the three measures informs a complex calculation which determines the quality-related research funding allocated from central government to UK HEIs.

At its most basic level many of the relevant dissemination metrics capture two key aspects of research.

1. The absolute volume (number) of research outputs. This simple count very clearly lends itself to a quantitative assessment of whether the FLF recipient has produced a higher number of research outputs than they did prior to receiving the FLF support. This is the 'within' person effect and considers the before and after count to identify any statistically identifiable change. It also lends itself to a quantitative assessment using a counterfactual approach.
2. The 'count' measure is informative and certainly establishes an individual as an active researcher in their field, however it leaves the question of the quality of the research open. The fundamental basis for academic research judgements is the peer review system. This system allows your 'peers' to make an assessment of the overall quality of each individual piece of work. This could range from an art historian expressing a view on the quality of the reasoning put forward by another on which of Van Gogh's paintings was his finest, to a computer scientist making a judgement about a short piece on artificial intelligence and pattern recognition on behalf of the Journal of Computer Science and Technology. The UK REF is based on peer review regardless of the nature of the output, be it a piece of sculpture or a published book.

Whilst there are many forms of output which differ substantively across disciplines, the most common is the peer reviewed journal article (although books, book chapters, and now some UK government sponsored reports are also subject to peer review). This has led to a proliferation of quality rankings of journals themselves in many subject areas. For example, in Business and Management the Chartered Association of Business Schools (CABS) produces a quality of journal ranking which encompasses a large share of the total peer reviewed journals (1,401 journals) in the 22 specific disciplinary fields. The ranking is reviewed approximately every three years and never within 12 months of a REF assessment year. Each journal is ranked on a scale of 4\* (the highest), 4, 3, 2, and 1 (the lowest). Indicative analysis of the last REF suggests an 80 per cent correlation between the CABS 2015 ranking and the final REF output ranking (which is on a scale of 4 = Quality that is world-leading in terms of originality, significance and rigour down to Unclassified = Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment).

However, given the fact that quality of journal, and implicitly the quality of the research that populates that journal, alone is no guarantee of the wider traction of an individual piece of research other metrics come into play. It is not uncommon for a 4\* ranked piece of research in a 4\* journal to have extremely low citation rates, despite the fact that many journal rankings include a citation count measure in their calculations. In this respect counting up the number of times an individual piece of research has been recognised by the community is an indicator of the quality, relevance and wider traction (or appeal) that a piece has on the assumption that it has shaped or influenced the new piece of citing research. This measurement lends itself to citation counts. Of course, citation counts are impacted by the absolute number of active researchers in your field (the potential pool of citers), the specific topic of the research paper, and other factors such as journal accessibility. This has led to the introduction of field weighted citation count metrics which compensate for some of these variations. In this sense, the recent developments in altmetrics (see below) and search engines that record views and downloads help capture wider engagement with research, and particularly from countries with limited access to journal subscriptions.

More recently, advances in the number of research-based search engines have led to the development of alternative metrics (Altmetrics). These advances have also been widely taken up by peer reviewed journals themselves. At the simplest level they often capture metrics relating to (a) how many people have read the abstract, (b) how many people have read the whole article, (c) how many people have downloaded the article, (d) how many people have tweeted a reference to the research, (e) how many shares and retweets the research has received, (f) how many likes, (g) how many recommendations and other social media based indicators of wider traction. What these new forms of metric allow is the capture of reach (how far your research has been dispersed) and traction (how widely it has been read, cited, and distributed). And they are particularly well-suited to capturing these two metrics amongst poorer individuals and countries where formal access to written forms of dissemination are too costly. Equally, it is possible to identify geographical reach on a country level.

To summarise, at the academic level there are a number of widely recognised metrics, search engines, and databases, which allow the identification of a set of outcome metrics that are appropriate for measuring and evaluating key aspects of research outputs and impact. This generally takes into account discipline specific differences in the nature of outputs and adjust for the specific nuances of each discipline. The vast majority of these metrics and systems for capturing them would be included in an individuals' case for promotion, new recruitment, and annual reviews. Whilst the FLF scheme, by definition, relates to 'leadership' *per se*, it is the case that in academia research leadership roles closely follow research performance in the sense that individuals who aspire to become leaders (for example Heads of Research at the discipline, faculty, college, or university level) must show excellence in their personal research and also some evidence of their ability to manage and co-ordinate research teams as well as develop strategies that deliver against commonly accepted research metrics. This follows the Research Development Framework (Vitae).

In the pure context of thought leadership, there is a very high correlation between seminal works, those that influence new thinking and shape and direct future research efforts and citations. Examples from business strategy would include Michael Porter's 5 forces research (Porter, M. E., 2008). This paper which has received 6,309 citations and has permeated into the business and academic world extensively. Equally seminal is the human genome research by a collective of researchers under the auspices of The International Human Gene Sequencing Consortium (International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921). The latter has received 5,751 citations.

**Table 5-3: Research dissemination based metrics**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Field-weighted citations at annual intervals	Academia	-	Both (ie STEM and AHSS)	Individual	Scopus	Counterfactual impact evaluation
Number of journal publications since start of fellowship -first author - contributing author	Academia	-	Both	Individual	ResearchFish Dimensions Scopus Fellow survey	Counterfactual impact evaluation Career tracker surveys
Number of publications since start of fellowship	Academia	-	Both	Individual	ResearchFish Dimensions Scopus Fellow survey	Counterfactual impact evaluation Career tracker surveys

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Average citation rate per publication	Academia	-	Both	Individual	Scopus	Counterfactual impact evaluation
Average number of publications per year	Academia	-	Both	Individual	Scopus Web of Science Google Scholar	Counterfactual impact evaluation
Average citation rate	Academia	-	Both	Individual	Scopus	Counterfactual impact evaluation
Field-Weighted Mass Media ratio	Academia	-	Both	Individual	Scopus	Counterfactual impact evaluation
Academic-Corporate Collaboration	Both	-	STEM	Individual	Scopus	Counterfactual impact evaluation
Field-Weighted Views Impact	Academia	-	Both	Individual	Scopus	Counterfactual impact evaluation
The Essential Science Indicators	Academia	-	STEM	Individual	Web of Science	Counterfactual impact evaluation
Substantial 'grey literature' outputs (policy reports)	Both	Both	AHSS	Individual	Fellow survey	Career tracker survey
Altmetric attention score	Academia	-	Both	Individual	Altmetric	Counterfactual impact evaluation

## 5.4 Networking, collaboration and influence metrics

Although we have identified measures of esteem and influence in terms of a research output which is most commonly in written form, and identified a general approach to capturing, measuring, and evaluating academic research performance we must not ignore other forms of output such as artistic & creative products (artefacts, images, artwork, score, creative writing, film/video/animation, exhibition, performance). These forms of output lend themselves to capturing wider forms of engagement and reach. For example, attendance at an exhibition or music recital or remote views and downloads. Databases such as Researchfish are important here as it captures a much wider set of outputs.



It is now appropriate to consider a broader set of metrics that capture elements of both research and leadership inputs and outputs including measures of wider esteem and influence. For example, **engagement activity** with working groups and expert panels, talks, presentations and debates, articles in magazines and newsletters, events and workshops, media interviews, press releases, blogs and other social media, and broadcasts. Also to consider measures that estimate reach and audience. Researchfish is a key source of data here, and also captures **perceived influence on policy, practice, patients and the wider public** including improvements in wellbeing and quality of life, and changes in efficiency and effectiveness of public service delivery

This has relevance and applicability to both academically led and produced research and that conducted within the context of an SME or large corporate. Both also have a direct relation to a wider leadership context where FLF supported individuals not only generate new knowledge and insights but co-ordinate the activities of others and engage with wider communities to produce, shape, and enhance the way things are ‘done’ in the future. Taking new innovations as a case in point, there are clear benefits to the originator from developing a new innovation and these benefits have a narrow personal (individual) return, but also a wider return (to the department and university or firm). These returns have a private and social element to them. But the largest societal gains are through the rapid dissemination, diffusion and adoption of new innovations. And it is this latter aspect that creates a wedge between the narrow private returns and the overall societal gains thus providing a justification for public support. Any evaluation of FLF must take into account all of these ‘layers’ when assessing outcomes and impact. At the heart of understanding and measuring these factors which relate to scale, reach and impact are the evolution of networks, the extent of formal and informal collaborations and metrics that capture wider influence and esteem (such as awards). Table 5.4 sets out a series of metrics that relate to these aspects of leadership and research evolution and influence and importantly have relevance to both academic and business pathways.

**Table 5-4: Measuring network breadth and scope (reach), collaboration, and wider outputs and influence**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
REF impact component for unit of assessment and institution - influence on policy and practice	Academia	-	Both	Department/cost centre Individual	Published on REF 2014 website (REF 2021, available in 2022) Researchfish	Counterfactual impact evaluation Theory-based impact (case study)
Number of other institutions collaborating	Both	Both	Both	Department	Fellow survey	Career tracker surveys

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
with at annual intervals						
Funding from more than one research council in time since start of fellowship	Academia	-	Both	Department	UKRI central data HESA income data Fellow survey	Counterfactual impact evaluation MI analysis Career tracker surveys
Impact metrics	Academia		Both	Individual	Researchgate	Counterfactual impact evaluation
Subject Area Count <sup>11</sup>	Academia	-	Both	Individual	Scopus	Counterfactual impact evaluation
Artistic/creative products	Both	Both	AHSS	Individual	Researchfish	Counterfactual impact evaluation
Any software/technical products developed	Both	Both	STEM	Individual	Researchfish	Counterfactual impact evaluation
Any datasets deposited with UK Data Archive	Both	Both	Both	Individual	Researchfish	Counterfactual impact evaluation
Any medical products, interventions or clinical trials	Both	Both	STEM	Individual	Researchfish	Counterfactual impact evaluation
Public events (public lectures, arts performance, exhibitions)	Academia	-	Both	Host	HE-BCI	Theory-based impact evaluation
Prizes	Both	Both	Both	Individual	Fellow survey	Career tracker surveys
Conferences	Both	Both	Both	Individual	Fellow survey	Career tracker surveys

<sup>11</sup> Subject Area Count using Scopus main or Sub-categories. "Publication-driven" assignment assumes that publications within a journal may have additional or different relevance to fields outside the core focus of the journal's scope. Publication-driven assignment offers the benefit of being able to assign individual publications from a journal separately to their relevant classifications.



Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Awards and Recognition <sup>12</sup>	Both	Both	Both	Host Individual	Researchfish Fellow survey	Counterfactual impact evaluation Career tracker surveys

## 5.5 Research funding metrics

Even the most basic research project requires a time commitment from a researcher or team of researchers. This is resource utilisation as people and their time are a finite resource. This is as relevant in academia as it is in business. It follows that for a given time resource (call it a working week for example) each individual can allocate their time to different activities. In academia we could broadly categorise these activities into teaching, research, and administrative duties. In business, it will often be the case that, at a moment in time, there are multiple products and services being developed, as well as multiple products and services being offered in multiple geographic markets. Even within a firm there are multiple teams and divisions to manage and coordinate. Each potential activity competes with every other activity for resources (here the time commitment of the individual). As such, if an individual is committed to one task then there is an opportunity cost associated with that choice which can be defined as the cost of not allocating time to all of the other potential tasks and roles.

We can also think of research funding as a financial investment in future research output or an investment in the leadership development of an individual. In this sense it is a (financial) resource input to the production of research, knowledge, and leadership capability. In its most basic sense, it buys the time commitment of the researcher or fellow in the case of FLF. As with any investment it has an expected return that exceeds that which would have been achieved if time had been allocated to another task. In financial markets we can think of the choice between putting a £1 in a savings account and receiving 0.5% interest on that £1 each year which is a safe option or investing that £1 on the stock market where it may increase or decrease in value (or stay the same). The expectation is that the FLF screening process is aimed at channelling funds to those individuals with the 'best' chance of achieving a positive return.

It is, however, often unrealistic to think of a single investment in isolation as becoming a thought leader or developing a new product or service is a process of discovery, testing, development and refinement and occurs over a period of time that extends beyond the

<sup>12</sup> Could include Research prize; Medal; Awarded honorary membership, or a fellowship, of a learned society; Appointed as the editor/advisor to a journal or book series; Poster/abstract prize; Attracted visiting staff or user to your research group; NIHR Senior Investigator/Clinical Excellence Award; National honour e.g. Order of Chivalry, OBE; Prestigious/honorary/advisory position to an external body; Personal invitation as keynote or other named speaker to a conference; Honorary Degree

initial investment in many cases. This has parallels in the world of risk capital investing where a venture capitalist will commit an initial investment across a portfolio of ventures and then provide follow-on rounds of investment to the most promising in that portfolio. We can also think of a similar process within a firm where it has multiple products and services in the development phase but only a subset of those (the most promising) receive a further resource commitment. This has parallels in academia where an initial body of research associated with a seed corn grant shows promise and this subsequently leads to further bids to explore exciting new avenues of research that build on the initial work. Table 5.5 identifies several metrics for capturing additionality in terms of an individual successfully securing further funding in an academic and business context.

**Table 5-5: Measuring additionality arising from initial funding**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Total research funding attained since start of fellowship/in addition to fellowship	Both	Both	Both	Department /cost centre Individual	HESA income data Scopus <sup>13</sup> Research fish Fellow survey	Counterfactual impact evaluation Theory based impact (case study) – looking for change in income trend Career tracker surveys
Number of new grants	Both	Both	Both	Department /cost centre Individual	As above	As above
Value of new grants	Both	Both	Both	Department /cost centre Individual	As above	As above
Source of new grants	Both	Both	Both	Department /cost centre Individual	As above	As above

## 5.6 Business performance metrics – businesses only

Ultimately business performance is typically considered in respect of ‘final’ outcomes such as share price appreciation, profitability, return-on-capital-employed (ROCE), firm value appreciation and other hard metrics. This reflects the dominance of large corporations and Tier 1 stock markets in the news media and also their relevance to individuals via their

<sup>13</sup> Awards Volume in SciVal refers to both the count and the value of grant awards. Grants are assigned to institutions based upon the Scopus affiliation of the principal investigator (PI) at the time of the grant being awarded.

pension funds which are often based on a portfolio of investments in 'blue chip' or FTSE 100 companies. It is also based on the concept of the shareholder who has a legal ownership stake in the firm and a residual claim on the future stream of profits of the investee firm.

However, these hard metrics are the final outcome of a process of investment in people, capital and technology, the development of new products and services, the expansion of market penetration and the penetration on entirely new markets. And this process is fundamentally shaped by the 'quality' of the management team and the strategic choices that they make. Below this strategic decision-making, superior outcomes can only be achieved when the core workforce has the resources and capabilities to fulfil these strategic objectives. In the case of innovation and technology this is often referred to as absorptive capacity (the ability to absorb and utilise new knowledge effectively and efficiently).

Economics tells us that, on average, being more efficient in terms of being able to turn inputs into outputs leads to firm and wider societal benefits including higher incomes and employment as resources are reallocated from inefficient firms to productive firms. In terms of how this occurs at the firm level, the three key components are people (and their embodied human capital), investment capital, and innovation. It follows that FLF may have a direct and an indirect influence on all of these three components in terms of the Fellow raising the human capital of the firm and quality of leadership, increasing the pool of financial resources for investment, and increasing innovation. In this sense, measures that capture this resource reallocation process (jobs and investment) through superior innovation and efficiency are relevant. These would include net job creation, income from patenting and licencing, which in turn would both reflect and add to economic efficiency through reducing the average costs of producing a desired (and higher) level of output. These metrics would be typical of a formal economic evaluation using a counterfactual approach. It would also allow estimates of deadweight and additionality which are central to an evaluation and CBA approach.

It is also the case that the 'classic' model of the corporation with its shareholders, board of directors, and salaried managers does not reflect the uniqueness of the smaller firm where ownership (shareholding) and control (management) is close and often resides with a single individual (often the founding entrepreneur) or her family. In the case of smaller firms' cash-flow based metrics may capture the essence of performance more appropriately as this is a fundamental determinant of risk of failure in the first instance and the ability to pay salaries and dividends to the owner(s) and management team. Retained cash is also correlated with the ability to invest in the future which in turn is associated with more formal performance metrics such as sales growth and efficiency.

Economic efficiency and productivity lie at the heart of any formal evaluation or measurement of outcomes associated with public interventions in the business sector. In essence increases in productivity, the efficiency with which a business turns factor inputs into outputs, is the single largest contributor to rising incomes and wealth at the macroeconomic level. There are two related concepts worthy of consideration here, both of which can be measured and identified at the firm level (or even below that for large

corporations with multiple plants producing similar outputs or even divisions). The first is technical efficiency which relates to the precise measurement of one's ability to turn inputs into outputs. This can be assessed at the individual level or at the level of the firm. The second efficiency measure is allocative efficiency which relates to how different factor resource inputs are combined to produce a mix of outputs. It follows that a superior decision-maker will (a) make superior allocative decisions, and, (b) given that resource allocation, ensure that technical efficiency is achieved. Table 5-6 identifies some relevant intermediate and final performance measures.

**Table 5-6: Measuring intermediate and final business performance**

Metric	Suitability			Unit	Data source <sup>14</sup>	Evaluation method
	Host	Business size	Discipline			
Company financial performance – profit as a % of turnover in most recent financial year	Business	Large	Both	Business	Annual Business Survey (ABS) Financial Analysis Made Easy (FAME) Business Structure Database (BSD)	Counterfactual impact evaluation Theory-based impact evaluation
Labour productivity – gross value added per full-time equivalent	Business	Large	Both	Business	ABS FAME BSD	Counterfactual impact evaluation Theory-based impact evaluation
% change in value of sales	Business	Both	Both	Business	IDBR ABS FAME BSD LSBS	Theory-based impact evaluation
Exporting (annual exports)	Business	Both	Both	Business	ABS LSBS SME Finance Monitor	Theory-based impact evaluation

<sup>14</sup> Whilst it may be possible to identify the company a fellow works for in each of these data sources, it is likely there will be too few fellows to conduct an impact evaluation. This is discussed in the impact evaluation chapter.

Metric	Suitability			Unit	Data source <sup>14</sup>	Evaluation method
	Host	Business size	Discipline			
Exporting intensity	Business	Both	Both	Business	FAME LSBS SME Finance Monitor	Theory-based impact evaluation
Cash-flows	Business	Both	Both	Business	FAME	Theory-based impact evaluation
% change in employment (or employment dedicated to R&I FTE)	Business	Both	Both	Business	IDBR ABS FAME BSD LSBS	Theory-based impact evaluation
% change in graduate employment	Business	Both	Both	Business	Case studies	Theory-based impact evaluation
Net new investment / Sales	Business	Both	Both	Business	ABS FAME	Theory-based impact evaluation
Expanded market position	Business	Both	Both	Business	Case studies Innovate UK	Theory-based impact evaluation
R&D expenditure	Business	Both	Both	Business	FAME	Theory-based impact evaluation
New product and service development	Business	Both	Both	Business	ABS FAME	Theory-based impact evaluation

## 5.7 Spin-offs/entrepreneurship metrics

Corporate venturing refers to the process by which a larger organisation (a business firm or university) becomes involved in the development, sponsorship, or investment of a start-up business in order to accelerate the development and commercialisation of new (often highly innovative) products or services. To facilitate this process many UK universities have a limited liability company that manages spin-off activity and some have a corporate venture fund to provide investment capital. In line with this Innovate UK have offered many

grant and funding pathways to support corporate venturing as a single or collaborative venture creation process with university generated knowledge and innovation at its core.

The process of corporate venturing lends itself to metric based measurement and we detail some core metrics to capture aspects of this process in Table 5.7. Due to its clear association with innovation and technical knowledge, corporate venturing is most closely aligned to STEM related disciplines where new knowledge is created and then becomes commercialised through a process of design, testing, and development of new products and services which are then commercialised.

**Table 5-7: Corporate venturing and spin-off activity**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Number of new start-ups/spin offs	Both	Both	STEM	Host Individual	ResearchFish (Spin Outs) <sup>15</sup> Financial Analysis Made Easy (FAME) Business Structure Database (BSD) HE-BCI Fellow survey	Counterfactual impact evaluation Theory-based impact Career tracker surveys
Any new research groups/teams established since start of fellowship	Both	Both	STEM	Host Department/ cost centre	Fellow survey Host survey	Career tracker surveys
Academic-Corporate Collaboration	Both	Both	STEM	Host Individual	Scopus <sup>16</sup> HE-BCI	Counterfactual impact evaluation
Academic-Corporate Collaboration income	Academia	-	STEM	Host	HE-BCI	Counterfactual impact evaluation

<sup>15</sup> ResearchFish collects name, registration number, date established, number of salaried employees, description

<sup>16</sup> Academic-Corporate Collaboration in SciVal indicates the degree of collaboration between academic and corporate affiliations: to what extent are this entity's publications co-authored across the academic and corporate, or industrial, sectors? A publication either exhibits academic-corporate collaboration, or it does not. This assignment is made based on the organization-type with which Scopus tags each affiliation.

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Academic-public/third sector Collaboration	Academia	-	STEM	Host	HE-BCI	Counterfactual impact evaluation
Academic-public/third sector Collaboration (income)	Academic	-	STEM	Host	HE-BCI	Counterfactual impact evaluation
Received venture funding	Both	Both	Both	Host Individual	CIS Beauhurst HE-BCI SME Finance Monitor	Counterfactual impact evaluation
Received follow-on funding rounds	Both	Both	Both	Host Individual	Beauhurst	Counterfactual impact evaluation
Developed a new product or service	Both	Both	Both	Host Individual	LSBS SME Finance Monitor CIS	Counterfactual impact evaluation
Commercialised a new product or service	Both	Both	Both	Host Individual	CIS	Counterfactual impact evaluation
IP and licencing	Both	Both	STEM	Host Individual	SME Finance Monitor	Counterfactual impact evaluation

## 5.8 Patent and IP activity metrics – STEM only

Patenting and licensing is not without contention, particularly in academia where concerns have been expressed about the potential for this form of legal protection to inhibit: freedom of expression and enquiry, publication of research findings, the diffusion of new knowledge; and also being beholden to outside ‘influence’. In the business (and academic) world patenting offers legal protection, which allows the creator (the inventor, entrepreneur, or firm) to exploit a (temporary) monopoly position and extract surplus rent (profit). Licensing allows firms to extract a fee per unit of sales derived. An interesting feature of the Covid-19 crisis was the relaxation of many relevant patents held by companies to ensure the widespread availability of key medical equipment (e.g ventilators) to the world.

Since the 1980s and 1990s, concerns over patenting and licensing in academia have diminished reflecting a broader pro-intellectual property protection culture and today many universities have a technology transfer office which manages the process of knowledge

transfer from academic patenting and licensing through to the development and commercialisation of academic research by private industry. This generally reflects the preference for universities for 'upstream' patenting activity which protects the insights and knowledge gained from base or primary (early stage) research that then feeds into the development of new techniques by industry which has a preference for patenting end products.

Park et al (2018) report a list of the 13 best known indicators:

1. Age of the patent;
2. Market value of the corporation applying for the patent;
3. Backward citations to the patent;
4. Forward citations to the patent;
5. Family size;
6. Scope (of applications using the patent);
7. Ownership;
8. Number of claims;
9. Strategy used for patenting;
10. Number of applicants;
11. Number of trans-border research cooperations involved;
12. Key inventors;
13. Legal disputes in regards to the patent (opposition in particular).

These and a number of broader metrics for capturing the essence of patenting and licensing activity are reported in Table 5.8. It also reflects the fact that the majority of patenting and licensing activity relates to medicine, science, engineering, computing and mathematical sciences.

**Table 5-8: Patenting and licensing metrics and measures**

Metric	Suitability			Unit	Data source <sup>17</sup>	Evaluation method
	Host	Business size	Discipline			
Any patents granted within X years of starting fellowship	Both	Both	STEM	Host Individual	CipherAI Espacenet Researchfish Dimensions HE-BCI	Counterfactual impact evaluation Career tracker surveys

<sup>17</sup> It is clear there are many sources which could be used to find patents each with their relative strengths and weaknesses. Cipher AI uses supervised machine learning in order to provide patent intelligence. Escapenet is a large database which would involve manual checking. There is also a suspected 18 month lag in data. Researchfish is based on the input of the researcher. Dimensions extracts data based on Researchfish. A fellow survey would be the quickest way to know if a patent has been granted to a fellow.



Metric	Suitability			Unit	Data source <sup>17</sup>	Evaluation method
	Host	Business size	Discipline			
Number of patents granted within X years of starting fellowship	Both	Both	STEM	Host Individual	Fellow survey (Short term) CipherAI Espacenet Researchfish Dimensions HE-BCI Fellow survey (Short term)	Counterfactual impact evaluation Career tracker surveys
Focus of patents aligned with Industrial Strategy	Both	Both	STEM	Host Individual	CipherAI Espacenet Researchfish Dimensions Fellow survey (Short term)	Counterfactual impact evaluation Career tracker surveys
Number of patents filed	Both	Both	STEM	Host Individual	Researchfish Dimensions <sup>18</sup> HE-BCI Fellow survey	Counterfactual impact evaluation Career tracker surveys
Extensions to IP	Both	Both	STEM	Host Individual	Researchfish Dimensions Fellow survey	Counterfactual impact evaluation Career tracker surveys

## 5.9 Leadership metrics

Leadership encompasses key elements of formal and informal human capital including the ability to organise work and resources, develop strategy, solve problems, and identify new opportunities. Being recognised as a ‘good’ leader has a personal aspect to it and a wider aspect to it. At the individual level the three primary leader activity dimensions include project management, personnel supervision, and strategic planning. Leadership quality is associated with being given more complex tasks, access to more resources, career enhancement both within and outside of the organisation, and all of these features should be associated with superior performance at the individual and host level. Internally, it is also associated with higher staff engagement and commitment to strategic aims. In a R&I environment wider measures of esteem (such as being invited onto a journal board, or an industry association board, or a government advisory panel) reflect the esteem an individual is held in by their wider peers be that within academia or business. In an academic setting the boundary between leadership and research esteem is particularly

<sup>18</sup> Dimensions has patent title, filing status (including application), legal status (eg granted, active, expired, pending or withdrawn), field of research, jurisdiction, ID and dates (filing date, publication date, expiry date); as well as supporting grants, inventor and organisation assigned to the patent

interesting, for example in terms of recruiting new PhD applicants but also managing them through to success completion of their studies.

An insight into what makes a good research leader, particularly in an academic context, is provided by the Leadership Lens on the Researcher Development Framework (Vitae, 2015<sup>19</sup>); and by the biennial Principal Investigators & Research Leaders Survey (PIRLS, now replaced by the Culture, Employment and Development in Academic Research Survey, CEDARS). PIRLS defines research leaders as those responsible for setting the intellectual direction of the research, holding research grants, and/or managing research staff or formally supervising postgraduate researchers. The 2015 survey (Mellors-Bourne and Metcalfe, 2015) finds that the key behaviours embodied by excellent research leaders are: advancing significantly their discipline/research area and exemplifying the highest standards of research integrity and conduct. Key leadership activities are building a research group and motivating individuals; although research leaders feel their institutions value research activity (research outputs including publications, securing research funding, collaborations outside HE, and academic collaborations) above leadership, management, impact and engagement activities.

**Table 5-9: Measuring leadership and influence**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
On journal editorial board, trade or industry body, or government advisory panel	Both	Both	Both	Individual	Researchfish Fellow survey	Counterfactual impact evaluation Career tracker surveys
On grant committees and/or assessment panels	Both	Both	Both	Individual	Researchfish Fellow survey	Counterfactual impact evaluation Career tracker surveys
Contribution to policy making (incl participation)	Both	Both	Both	Individual	Researchfish Fellow survey	Counterfactual impact evaluation

<sup>19</sup> This sets out the knowledge, behaviours, attributes and skills researchers are expected to develop during their careers, and is an evolution of the Research Councils' Joint Skills Statement. It provides a tool for individuals and researcher developers to plan and evaluate professional development; and when using the Leadership Lens draws out the behaviours and attitudes that constitute good leadership. The RDF has 4 domains: Knowledge and intellectual abilities, Personal effectiveness, Research governance and organisation, and Engagement, influence and impact; and 12 sub-domains. The Leadership Lens identifies behaviours and attitudes across these. <https://www.vitae.ac.uk/vitae-publications/rdf-related/leadership-lens-on-the-vitae-researcher-development-framework-rdf-april-2012.pdf/view>

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
in national consultation)						Career tracker surveys
Number of PhDs supervised in past year	Academia	-	Both	Host Individual	Fellow survey PIRLS/CEDARS	Career tracker surveys
Number of PhDs examined in past year	Academia	-	Both	Individual	Fellow survey	Career tracker surveys
Number of research staff managed (excl PhD)	Academia	-	Both	Host Individual	Fellow survey PIRLS/CEDARS	Career tracker surveys
Develop research staff	Both	Both	Both	Host Individual	Fellow survey PIRLS/CEDARS	Career tracker surveys
Supported/advised grant applications of others	Both	Both	Both	Host Individual	Fellow survey PIRLS/CEDARS	Career tracker surveys
Academic collaboration/partnership -within discipline -across disciplines -within UK - international	Both	Both	Both	Host (academia only) Individual	Fellow survey PIRLS/CECARS (academia only) Researchfish	Career tracker surveys
Collaboration/partnership outside HE	Both	Both	Both	Host (academia only) Individual	Fellow survey PIRLS/CECARS/HE-BCI (academia only) Research fish	Career tracker surveys
Staff satisfaction, engagement and commitment	Both	Both	Both	Host Individual	Fellow survey CROS/CEDARS (academia only)	Counterfactual impact evaluation

## 5.10 R&I metrics and wider reach

Successful leadership and research (as with seed-corn risk capital investments) often lead to three tangible intermediate and final outcomes:

1. More internal resources are allocated to leaders and research teams that demonstrate high performance and future potential.
2. It also becomes easier to leverage external resources.
3. Leaders, research projects and teams who get wider esteem and recognition have an enhanced probability that external people and organisations want to collaborate with them.

The latter two have parallels in equity issuance by corporations where syndicated underwriting (a shared commitment to purchase the shares) is common as indeed in syndicated venture capital deals (shared investing) which often follow the two heads is better than one approach to gauging the quality of a deal or investment project.

In Table 5.10, we include metrics that capture elements of internal resource accumulation such as managing a larger team, a shift towards more permanence in the research teams contracts of employment (a longer-term commitment), and the degree to which the whole team is able to achieve career progression. We can also measure broader research performance elements in academia through REF outcome metrics and over REF cycles identify changes in staff entered and outcome metrics over time at the Unit of Assessment (which roughly maps into disciplines) level. At the business level we can identify similar resourced based indicators such total employees in the team and the wage bill associated with them, number of staff engaged in R&D etc. For both academic and business pathways we can capture elements of wider esteem, relevance and traction through the number and scale of external collaborators and collaborative activities. Some of these overlap with leadership metrics noted above, so could be included within this theme or within the leadership theme.

As one of the FLF goals is to enhance inclusion and diversity, we also identify and relevant measures to capture team development in respect of ethnicity, non-nationals, and gender and its evolution over time.

**Table 5-10: Research & Innovation scale and scope metrics**

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
% change in number of employees within X years of starting fellowship	Business	SME	Both	Host	ABS FAME BSD	Counterfactual impact evaluation

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
% change in number of employees within department within X years of starting fellowship	Academia	-	Both	Department /cost centre	HESA Staff Survey Fellow survey	Counterfactual impact evaluation Career tracker surveys
% of researchers on open-ended contracts observed at annual intervals	Both	Both	Both	Host Department /cost centre	HESA Staff Survey (academia only)	Counterfactual impact evaluation
Proportion of fellows staying in research, observed at annual intervals	Both	Both	Both	Programme Host Individual	MI Fellow survey	Theory-based impact evaluation Career tracker surveys
Moving employer (within or across sectors)	Both	Both	Both	Individual	Fellow survey	Career tracker surveys
Moving location (beyond UK)	Both	Both	Both	Individual	Fellow survey	Career tracker surveys
Overall REF submission scores for unit or assessment/institution weighted by FTEs	Academia	-	Both	Host	Published on REF 2014 website (2022 for 2021 results)	Counterfactual impact evaluation Theory-based impact evaluation (case studies)
Size of research team observed at annual intervals	Both	Both	Both	Department	Fellow survey	Career tracker surveys
Whether junior staff are being promoted	Both	Both	Both	Department	Fellow survey	Career tracker surveys
Number of higher degrees registered in past year	Academia	-	Both	Host Department /cost centre	HESA Fellow survey	Career tracker surveys
Number of other institutions collaborating with at annual intervals	Both	Both	Both	Department/ individual	Fellow survey	Career tracker surveys

Metric	Suitability			Unit	Data source	Evaluation method
	Host	Business size	Discipline			
Funding from more than one research council in time since start of fellowship	Academia	-	Both	Department/ cost centre	UKRI central data HESA income data Fellow survey	Counterfactual impact evaluation MI analysis Career tracker surveys
Any new research groups established since start of fellowship	Both	Both	Both	Department	Fellow survey	Career tracker surveys
Total wage bill of direct reporting staff pa	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Number of direct reporting staff pa	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Total staff training and development budget pa	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Total number of staff receiving training and development pa	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Proportion of staff from ethnic minority groups	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Proportion of staff from outside UK	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys
Proportion of female staff	Both	Both	Both	Department, Host	Fellow survey	Career tracker surveys

## 5.11 The timing of outcome metrics

In this section we consider what the relevant time period or timing is for each set of metrics across our eight thematic strands. This is important in evaluations as not all potential outcomes will have occurred in an early evaluation and many of the relevant metrics in an early-stage evaluation have less relevance to a later stage evaluation. It may also be the case that the ability to measure something is constrained by the ability to capture the relevant data (for example the UK REF is only conducted every five to six years) or that there is a lag between an event occurring and the administrative recording of that event (for example patent filing through to the patent being granted). Equally, many organisations make three-to-five years' strategic plans that shape their resource allocations for that period. Exogenous events, such as Covid-19, may alter resource allocations and lead to delay and suspension of investments that would have occurred. The 'J' curve effect is a

well-established feature of disruptive change and radical innovation investment. In brief, in the short-term after a disruptive or radical change (for example an FLF award), performance on standard performance metrics may dip as people, resources and systems are reconfigured and aligned to the new strategy. Thus, comparing against what seemed like a suitable comparator in the short-term may confuse the evaluator into concluding that the scheme impact is poor. This is the bottom of the 'J'. Subsequently, as we rise up the 'J' performance may improve in an exponential way.

**Table 5-11: Mapping Metrics to the position in TOC**

Timing of measurement				
Theme and metrics	Short-term 0-3 years	Medium-term 3-5 years	Long-term 6 years plus	Position in TOC
<b>Research</b>				
Quantity of outputs	✓	✓		<b>Fellow outcomes</b> – related to becoming outstanding researchers
Quality of outputs		✓	✓	
Altmetrics	✓	✓	✓	<b>R&amp;I outcomes</b> – related to initially ambitious R&I and later leading edge R&I
<b>Networking, collaboration, influence, leadership</b>				
Networking	✓	✓	✓	<b>EDIP pathway</b> - FLF builds the conditions for increasing collaboration, positive competition
Collaboration		✓	✓	
Influence		✓	✓	
Supervision	✓	✓		<b>Fellow pathway</b> – becoming an influence, international research leader
<b>Research funding</b>				
Volume of new projects/grants secured	✓	✓	✓	<b>Fellow pathway</b> – becoming an influence, international research leader
Value of new projects/grants secured	✓	✓	✓	
<b>Business performance</b>				
<b>Input side</b>				
Employment change	✓	✓		<b>Fellow pathway</b> – certainty from host commitments in employment T&Cs
Staff development	✓	✓		
New Investment	✓	✓		
R&D spend	✓	✓	✓	<b>Host pathway</b> – funding builds commitment
New products & services		✓	✓	<b>R&amp;I</b> – ambitious/leading edge R&I
<b>Output side</b>				
Hard financial measures		✓	✓	<b>Host</b> – sustained in R&I
Productivity / efficiency		✓	✓	



Internationalisation		✓	✓	
Corporate venturing				
Venturing activity count		✓	✓	<b>Host</b> – sustained in R&I <b>EDIP pathway</b> - FLF builds the conditions for increasing collaboration, positive competition <b>R&amp;I</b> – ambitious/leading edge R&I
Collaborations	✓	✓	✓	
Funding		✓	✓	
New product / service development	✓	✓		
New product / service commercialisation		✓	✓	
Patenting				
Filed applications	✓	✓		<b>R&amp;I</b> – ambitious/leading edge R&I
Granted		✓	✓	
Extended			✓	
Research & Innovation scale and reach				
Team size	✓	✓	✓	<b>Fellow</b> – influential, international leader
Team diversity		✓	✓	
Employment contracts	✓	✓	✓	<b>Host</b> – sustained in and committed to R&I
Retention			✓	
REF outcomes		✓	✓	
Spatial reach		✓	✓	
Institutional reach		✓	✓	

In essence, Table 5.11 maps out the processes by which core funding shapes the career development of the individual, then the context of the local environment they operate in, the wider organisational impact and their external reach and influence. For example, an FLF fellow may initially seek to build a research team around them be it in an academic or corporate environment. This team may engage in foundational thinking that manifests itself in a tangible way through more innovative research output which then translates into wider measures of quality and esteem and draws in more individual and organisational networks and collaborations over a period of time. At the level of a product or service, this is underpinned by new thinking and insight, which translates into core R&D around a new product or service which the organisation might seek to commercialise and patent or license. The ability to innovate then translates into a demand for more factor inputs (jobs, capital investment, etc) as an innovation creates new demand and the potential to internationalise etc.

## 6 Considerations for a counterfactual impact evaluation

This chapter sets out a proposed approach to estimating the causal impact of FLF. It builds on the Theory of Change (Chapter 2) and explication of the challenges associated with evaluating FLF (Chapter 4).

The first part of the chapter reviews the scope and options for estimating causal impacts. It explains the basic intuition behind estimating impact and details the information available from FLF applicants and the FLF scheme and selection process to inform this exercise. It then explores options for comparison groups and the advantages and disadvantages of each, including likely availability of data.

The second part of the chapter sets out the recommended approach to estimating the impact of FLF. We explain the proposed methods, the assumptions which must be met if they are to provide a robust estimate of impact and why they are likely to be appropriate in the case of FLF. It also sets out the limitations of the analyses and the extent to which the chosen approaches are likely to be able to overcome the challenges identified in Chapter 4. It also explains why this approach has been recommended over other common methods: Box 4 provides a summary of other methods commonly used to estimate the causal impact of policy interventions and explains why they would not be suited to estimating the impact of FLF.

The proposed approach to data collection is outlined, along with details of the other options considered.

### Key findings for an evaluation design for FLF:

- The main focus to estimating the causal impact of the FLF programme should be on estimating the average impact of FLF on those offered a fellowship across the range of outcome measures which can be observed in secondary data sources. Using bespoke surveys to measure outcomes for fellows and a comparison group is not recommended due to likely imbalance in response rates and attrition.
- The comparison group would be drawn from unsuccessful applicants using baseline data collected during the application process, and the estimate of impact obtained through Propensity Score Matching (PSM). Other methods for identifying causal impact (Randomised Control Trials, Difference-in-Differences, and Survival Analysis) are not suitable for an evaluation of FLF.
- The analysis should explore the sensitivity of the impact estimates to varying the choice of matching estimator and the closeness of the match between the treatment and comparison groups, including limiting the potential comparators to individuals who were invited to interview.

- The analysis should focus primarily on outcomes and impacts at the level of the individual, as although outcome metrics may exist to estimate the impact of FLF at different levels (research teams, departments, host institutions, external collaborators, R&I ecosystem and wider society) it may not be credible to attribute any differences in observed outcomes between the treatment and comparison groups which emerge over time to the impact of FLF alone.
- Identifying the comparison group and tracking both fellows and the comparison group in secondary data sources requires permissions to process data, this is governed by the Je-S.
- If additional resource was available and data sharing agreements in place, this approach could be supplemented with an analysis of the effectiveness of FLF compared to other fellowship programmes (notably Sir Henry Dale Fellowships and ECR Starting Grants for some segments of FLF). However this relies on obtaining access to similar baseline data on participants in other fellowship schemes, as well as the personal data required to track their outcomes over time. This analysis would also use PSM.
- Depending on sample sizes, it may be informative to use a Regression Discontinuity Design (RDD) to assess whether FLF has an impact at the margins of treatment. This would provide evidence on whether the impact of FLF was also apparent when varying the approach to the analysis. Ideally this analysis would be restricted to the subset of applicants who were interviewed for FLF, but it is likely to be necessary to include the wider pool of applicants around the margins of treatment based on sift scores to increase the likelihood of sample sizes being adequate to be able to detect any impact.

## 6.1 Estimating causal impact

To estimate the causal impact of FLF it is necessary to form an estimate of the outcomes that those offered a fellowship (known as **the treatment group**) would have attained if they had not received the funding. This is known as **the counterfactual**. The estimate of counterfactual outcomes is compared against observed outcomes for fellows and the difference in outcomes can be attributed to the impact of FLF.

Most approaches to identifying causal impact involve observing outcomes for a **comparison group** and using statistical methods to adjust these for any differences in the characteristics of the treatment and comparison groups which might affect the outcomes that either group attains. The use of counterfactual impact evaluation methods improves the likelihood that the estimate of impact takes account of any changes in outcomes which would have occurred over time even without the intervention. For example, early career innovators and researchers may publish their work and achieve career progression over time even without FLF, but changes in the research funding environment may have an impact on their ability to do so. Failing to take into account the trajectory in outcomes which is likely to occur naturally over time, as well as changes in the wider research environment, is likely to mean that the impact of FLF is over- or under-estimated.

### 6.1.1 The selection process and application data

Those applying for the FLF scheme are asked to provide:

- personal data, such as their name and gender,

- 
- details of their background characteristics, such as ethnicity, nationality, age at the date of application, whether they had a PhD and if so, when it was completed,
  - information on their employment at the time of applying for the fellowship, including the name and region of their host organisation and whether they were employed on a permanent contract,
  - the percentage of time (as a full-time equivalent) that they would expect to spend on the fellowship,
  - other details of the support requested, including the amount, whether they had project partners and the contribution of those partners in terms of cash or other in-kind support.

The applications database also records the round the application was made under, the number of external reviewers, the mean peer review score, the number of councils involved in considering the application, the lead council (which gives the broad discipline of the Fellow and the research, more detailed discipline or field is not recorded), the sift score and the interview score (just for those considered suitable for interview). The stage at which the application was rejected, and whether the candidate was invited to interview, is also recorded.

In Round 4 there were 14 different selection panels for the FLF scheme, with some roaming members providing consistency between panels (although scores do nonetheless appear to differ by broad discipline). Peer review scores are discussed and calibrated by the panel and each candidate is given a consensus sift score out of 10. Based on this, a decision is made on which candidates to invite for interview. Applications are banded between A and D to indicate the relative priority for funding. Those identified as a high priority are selected first if the number of applications of sufficient quality to fund exceeds the number of awards available in a given round (see also Chapter 3 for details of the application process and outcomes including distribution of scores by panel/broad discipline).

### 6.1.2 Potential comparison groups

This section considers possible comparison groups which could be used to estimate the counterfactual and explains how the choice of comparison group affects the research questions that can be answered. It also sets out data requirements and assesses the feasibility of obtaining the information required to use each comparison group to estimate the impact of FLF.

As noted in Chapter 2 on the Theory of Change, impacts from FLF are expected to be seen at a range of different levels, including the individual fellow, the host organisation, across the research and innovation community and more widely at UK and international level. Whilst the individual is the focus in identifying potential comparison groups, it would nevertheless be possible to observe outcomes at higher levels of aggregation e.g. for host organisations. This issue is discussed further in section 6.2.4.

## Unsuccessful applicants

In the first four rounds of applications FLF places were heavily over-subscribed: 305 fellowships were awarded to 1,553 applicants, suggesting that there were around five applicants for every fellowship. The ratio of applicants to places was much higher in the business sector (nine applicants to every place), but even in academia there were more than four applicants to every fellowship awarded. The ratio of applicants to fellowships therefore suggests that there is a good prospect of identifying a well-matched comparison group from the pool of unsuccessful applicants.

The information supplied as part of the application process provides rich data on both successful and unsuccessful candidates in terms of their background characteristics, the support requested and their likely suitability for a fellowship. This means that it is likely to be possible to identify a comparison group from the pool of unsuccessful applicants who are similar to successful applicants across a wide range of characteristics. Having access to rich data on a potential comparison group observed at the same point in time as the treatment group increases the likelihood of being able to identify comparators likely to experience similar outcomes to the treatment group from the pool of unsuccessful applicants.<sup>20</sup>

There are two main advantages to drawing the comparison group from the pool of unsuccessful applicants. Firstly, individuals who apply for the FLF scheme may be more focused on achieving long-term career goals and taking a leadership role than the wider pool of early career researchers and innovators. If this is the case, there may be a difference in the outcomes achieved by applicants and non-applicants. A comparison group drawn from unsuccessful applicants is likely to provide a more credible estimate of the counterfactual than if the potential link between motivation to apply for the fellowship and motivation to achieve the types of outcomes expected from participation in the FLF scheme was ignored.

Secondly, all applicants for the FLF programme supply personal data as part of the application process. Provided it is possible to use these data on unsuccessful applicants in the evaluation, this creates the opportunity to track their outcomes in secondary data sources over time and observe the same outcomes for both fellows and this comparison group. Having access to background characteristics and outcomes for unsuccessful applicants offers a substantial advantage over what is likely to be possible for other potential comparators.

Using unsuccessful applicants to form the control group is an approach used in several other evaluations of fellowship programmes and is the preferred approach outlined in the Innovate UK evaluation framework (see Chapter 4).

---

<sup>20</sup> Note that most matching approaches would not use all unsuccessful applicants as the comparison group. Rather, the comparison group would be based on the closest matches and weight individuals according to the closeness of the match.

### Applicants for other Fellowships

A potential difficulty in drawing the comparison group from unsuccessful applicants for the FLF scheme is that at least some of the unsuccessful candidates are likely to be a poor match for those who are offered a fellowship. Unsuccessful candidates may be of lower quality than those offered a fellowship and as a result they may be less likely to achieve the same outcomes. If the pool of unsuccessful candidates contains few individuals who are of the same quality as those offered a fellowship, this could result in an upward bias in the estimated impact of the FLF scheme.

Drawing the comparison group from a pool of individuals who have been offered places on alternative fellowship programmes may reduce the likelihood of over-estimating the impact of FLF. This is because those awarded fellowships under other schemes might be of higher quality than unsuccessful applicants for the FLF programme. As a result, they may be more likely to experience similar outcomes to those awarded a FLF. However, there are a number of reasons why it is unlikely to be advisable, or feasible, to base the estimate of the impact of FLF on a comparison with the outcomes experienced by those participating in other fellowship schemes:

- Other programmes will have different aims to FLF and may appeal to or be restricted to applicants with different characteristics and ambitions (see Chapter 3). This is likely to affect the outcomes that participants in other fellowship programmes experience. It is difficult to identify other fellowship schemes with similar aims and likely to attract a similar range of candidates, given that FLF is open to researchers and innovators across both academia and business and from all disciplines.
- It would be necessary to limit the comparison group to those being selected for other fellowships at the time the selection process for FLF is running, to ensure that outcomes for either group are similarly affected by other macroeconomic changes likely to affect outcomes over time. This may restrict the numbers of potential comparators available.
- With fellowship schemes running concurrently, the comparison group will have had the option of applying for FLF but instead choose to apply for an alternative scheme (as for many schemes you may only apply for one). This raises the question of whether applicants for other fellowship schemes are actually similar to those participating in FLF and whether they would be likely to experience similar outcomes, given that they have opted to apply for a different type of fellowship.
- If the comparison group is drawn from those participating in other fellowships, they will receive other types of funding and support with the intention that this will affect the outcomes that particular scheme seeks to influence. It is likely that some of these outcomes will overlap with the expected outcomes from participation in FLF. This means that the evaluation of FLF will capture its impact relative to other fellowship schemes, rather than its impact compared with not participating in FLF. This reduces the likelihood of detecting any impact from FLF.
- To estimate the relative effectiveness of FLF compared with other fellowship programmes it will be necessary to have access to similar baseline data on those awarded other types of fellowships as that available on those participating in FLF, as

well as the ability to track their outcomes over a similar period of time. This will only be possible with access to personal data on participants in other fellowship schemes and permission to track their outcomes. Some of these data (for example name and host) may be available on funders' websites as they tend to publish the names of successful awardees. However, the feasibility of the analysis rests on whether it is possible to gain access to data on individuals participating in fellowships across the same full range of disciplines to be included in the impact evaluation of FLF.

Given the reduced likelihood of detecting any impact from FLF when outcomes for FLF participants are compared with outcomes for those participating in other fellowship schemes, as well as the potential difficulties in gaining access to data on a comparison group who are similar to FLF participants, there is likely to be limited value in using participants in other fellowship programmes as a comparison group. However, if resources allow, obtaining data on this comparison group (particularly Sir Henry Dale Fellows for academic STEM FLF Fellows and ERC Starting Grant Fellows for FLF academic fellows, as indicated in Chapter 3) those would address the question of whether FLF has a greater impact on particular outcomes than other types of fellowship programme.

### Early Career Researchers and Innovators

A final potential comparison group might be drawn from the wider pool of early career researchers and innovators. The difficulties inherent in identifying individuals who would be offered an FLF from those who have not actually applied for FLF, or any other type of fellowship, makes this unlikely to be feasible however. Even identifying early career researchers and innovators is likely to be difficult, as there is a wide age range amongst applicants for FLF (including several applicants over the age of 50). Whilst it is more common for applicants to be in their early 30s, it may be difficult to find close matches for some individuals in receipt of FLF if the comparison group is restricted to a narrower age range. Also, there is no single individual-level dataset providing: a census of all early career researchers; that can provide rich information on circumstances around the time that the FLF programme started; and spans both academia and business. This would be needed to make it possible to identify early career researchers and innovators based on employment history, rather than age. For example, Scopus and Researchfish do have good coverage but there are issues with discipline bias and quality; the HESA staff record identifies and captures data on early career researchers in academia (under a contract of employment) but not those working in business; and the Postgraduate Research Experience Survey (PRES) captures data including motivation, professional development and future plans for those on a doctoral programme but not their actual employment beyond the programme (see Appendix E).

Those who choose not to apply for FLF may have different characteristics to those who do apply. For example, the extent to which academic institutions encourage and support their staff to apply for fellowships is likely to vary. Applicants for FLF may also be more focused on the types of outcomes that the programme seeks to promote than early career researchers and innovators more generally. With access to detailed individual-level data on a large group of early career researchers it may be possible to identify a comparison group with similar observed characteristics to FLF participants but would be more difficult to ensure that the two groups were well-matched on unobserved characteristics which might



be related to outcomes, such as motivation. If the comparison group are less motivated than FLF participants, the impact of FLF may be overestimated.

It would be necessary to have access to personal data on early career researchers and innovators, as well as detailed baseline information similar to that observed for FLF participants to track outcomes for this comparison group over time. It seems unlikely to be possible to gain access to personally identified information for the full cohort of early career researchers and innovators who would be eligible for FLF. As there are also reasons to believe that any estimate of impact which uses early career researchers and innovators as the comparison group would not provide a credible estimate of the impact of FLF, there is little value in pursuing this option.

### **6.1.3 Feasibility of processing personal data for the counterfactual impact evaluation**

The evaluation will need to process personal data to: a) track fellows against a comparison group (recommended as unsuccessful applicants); and b) support the matching process to identify a suitable comparison group (and increase the likelihood of achieving a match for the majority of cohorts). The key approach suggested is to use programme monitoring information and secondary, administrative data sources to match and track individuals for a counterfactual impact evaluation; so there would be no requirement to contact individuals for survey or qualitative interviews in the counterfactual impact evaluation. Ideally unique identifiers would be used, although this relies on the same identifiers being used by UKRI and the administrative sources (e.g ORCID). However, whether or not the same identifier is used, data may be mis-entered or contain errors, hence using personal data to enable for fuzzy matching increases the prospects for matching a higher proportion of fellows. Consequently, it is important that permissions allow personal data processing for the purposes of research and/or evaluation.

The permissions held by UKRI on applicants and grantees are therefore crucial to consider as part of the feasibility study. The terms and conditions<sup>21</sup> that cover FLF applicants are those of the JeS (Joint Electronic System) which is the on-line electronic system used by seven Research Councils (AHRC, BBSRC, EPSRC, ESRC, MRC, NERC and STFC), to receive funding applications.

The data controller needs to reach decisions about the potential to release personal data for processing in relation to tracking in administrative data. It might be judged that the conditions are in place to allow personal data to be used for tracking in administrative data-sets. Either way, the commitment made in the Je-S is that individuals will be notified of any changes to the terms and conditions - which suggests there is potential for UKRI to clarify how it would use personal data in the FLF evaluation.

---

<sup>21</sup> <https://je-s.rcuk.ac.uk/Jes2WebLoginSite/TermsConditions.aspx?mode=accountsetup>

### 6.1.4 Summary

Drawing the comparison group from the pool of unsuccessful applicants for FLF is likely to provide the most viable way of estimating the impact of FLF on those meeting the eligibility criteria to be offered a fellowship. There is a good chance of being able to use the data collected in the application process to observe baseline characteristics and to track outcomes for both the treatment and comparison groups, provided it is possible to use personal data to track both groups in secondary data sources. This is also likely to provide the most robust estimate of the impact of FLF, as it is possible to use the information collected at baseline to ensure that the comparison group is drawn from unsuccessful applicants who have similar characteristics to successful applicants. There are also grounds to believe that unsuccessful applicants for FLF are similar to those who are offered a fellowship in terms of motivation – something that is likely to have an impact on their subsequent outcomes. Observed outcomes for a well-matched group of unsuccessful FLF applicants are therefore likely to provide a credible estimate of the outcomes that FLF participants would have attained if they had not received funding from the FLF. This approach would also have the strongest likelihood of detecting any impact from FLF.

Should budgets and data sharing arrangements allow, it may also be informative to estimate the impact of FLF relative to other fellowship schemes. This would require access to personal data, baseline characteristics and outcomes for a comparison group of participants in other fellowship schemes, and would only be possible if this is available. It would potentially demonstrate whether FLF was more or less effective than other fellowship schemes, but there would be a lower likelihood of detecting any differences in impact due to the fact that the comparison group would receive support from another type of fellowship. It is therefore possible that the analysis would be inconclusive and so it is recommended that this analysis be a lower priority for funding.

## 6.2 Proposed approach to estimating the impact of FLF

### 6.2.1 Propensity Score Matching

Propensity Score Matching (PSM) seeks to estimate the impact of an intervention by comparing outcomes for individuals in the treatment and comparison groups with a similar propensity to take part. For this approach to provide a robust estimate of impact it is necessary to correctly identify the characteristics that determine both the likelihood of participating in FLF and the outcomes that result. This requires access to detailed and accurate information on the treatment and comparison groups before and after the intervention.

It seems likely that the data collected as part of the application process, including personal data which can be used to track outcomes for applicants over time, would be adequate to use PSM to estimate the impact of FLF, although there are some limitations. For example, the data only records the UKRI councils involved in considering the application, rather than more detailed information on disciplines. Information on where the individual is located is also at a high level of aggregation, although potentially it may be possible to use the details

of the employing organisation to add supplementary information on the local area from secondary data sources.

Rather than matching individuals in the treatment and comparison groups on particular characteristics, PSM matches them on a score derived from many different determinants of participation. The first step is to estimate the probability of an individual with a given set of characteristics being offered a fellowship. It is then possible to compare outcomes for individuals with a similar propensity to be treated, but where some are given FLF and others are not. It is unlikely to be possible to match treated individuals to untreated individuals with an identical propensity score, so instead PSM selects matches within a certain range of the propensity score for the treated individual.

The extent to which the characteristics of the treatment and matched comparison groups are similar after matching (known as the balance) is assessed to gauge the likelihood that the impact estimates are biased by differences in the characteristics of the two groups which remain after matching. If the treatment and matched comparison groups appear similar across a wide range of observed characteristics after matching it is more credible to believe that any differences in outcomes between the two groups are due to the impact of FLF, rather than differences in the composition of the treatment and comparison groups.

The closeness of the match between the treatment and comparison groups determines whether PSM provides a reliable estimate of the impact of FLF. If it is not possible to find close comparators for all participants, the impact estimate is unlikely to be representative of the impact of FLF across all fellows. It will be important to assess the sensitivity of the impact estimates to varying the closeness of the match between the two groups and to report the percentage of FLF participants who can be matched to members of the comparison group with similar propensity scores to assess both the robustness and the generalisability of the findings. It is also usual to explore the extent to which the impact estimates are sensitive to using different types of matching estimator. Depending on sample sizes, it may also be worthwhile to consider how restricting the matched comparison group to a more limited sample of unsuccessful applicants affects the impact estimates. For example, limiting potential comparators to individuals who were selected for an interview for FLF may improve the credibility of the analysis.

Consideration of potential outcome measures (see Chapter 5) finds that some outcomes are only relevant to STEM or AHSS subjects, or to academic or business settings. This suggests the need to segment the population into four groups: STEM academic, AHSS academic, SME business and large business. The matching of treatment and comparison groups should be constrained to those within the same segments. Analysis of programme MI from Rounds 1 to 4 (see Chapter 3) finds these groups are vastly different in size. This may mean that the counterfactual impact evaluation can only viably focus on academic settings as the numbers of applicants and fellows from business, and large firms in particular too small to support analyses. Theory-based impact evaluation methods will need to be employed (eg matched case studies) for the smaller business segments (see Chapter 7).

PSM would also be suited to estimating the relative impact of FLF compared with other fellowship schemes if it is possible to obtain data on sufficient numbers of well-matched

comparators and a similar range of baseline characteristics to those collected in the FLF application process.

### 6.2.2 Regression Discontinuity Design

As mentioned earlier, applicants for FLF are scored as part of the peer review process and are then given a sift score by the panel which reviews their application. The way in which candidates are assessed is specific to a particular subject or discipline, but the panels include some roving members to ensure that the scores given by the different panels are comparable. A subset of candidates are selected for interview based on the results of the sift process, and they are also scored based on their performance at interview.

As all applicants for FLF are given a sift score, and interviewed candidates are scored on their performance at interview, this creates the opportunity to identify the subset of candidates who were just over the minimum threshold to be offered a fellowship and those who just missed out on a fellowship. A regression discontinuity design (RDD) compares outcomes for those either side of this threshold to produce an estimate of the impact of FLF at the margins of treatment. It is likely that those around this threshold will be similar and so it is credible to assume that any difference in outcomes between the two groups will be largely due to the fact that one group receive the fellowship whilst the other group does not.

A limitation of a RDD is that it only provides an estimate of the impact of FLF on those at the margins of treatment, rather than averaged across the wider pool of fellows. If those who are only just selected for FLF differ from the wider pool of fellows, the impact estimate may not be representative of the impact across all FLF fellows.

RDDs are considered 'sharp' if there is a hard cut-off which determines whether an individual receives the treatment. For example, an intervention may only be available to those under the age of 25. Those aged 24 and 25 would be at the margins of treatment, with those aged 24 eligible and those aged 25 ineligible. In the case of FLF the threshold which determines treatment is 'fuzzy', as the decision on which individuals should be offered the fellowship depends on their performance at interview, as well as the sift score. This means that even amongst individuals with the same sift score, some may be given a fellowship whilst others may not.

Whilst the treatment and comparison groups are likely to be better matched if the RDD focuses on those interviewed for FLF, rather than those with sift scores around the observed minimum for those offered a fellowship, as only 550 fellowships are due to be awarded in total, the numbers of interviewed individuals around the margins of treatment may be insufficient to detect any impact from FLF.<sup>22</sup> It may therefore be necessary to include all individuals with similar sift scores to successful applicants who just scored highly enough to be selected to implement an RDD. In most rounds the lowest sift score for a successful candidate was 7, although in Round 3 a candidate with a sift score of 6 was awarded a fellowship. The vast majority of candidates with a sift score of 8 or above were invited to interview, but only a small proportion of those with a score of 6 or 7 were offered

---

<sup>22</sup> Around half of those interviewed were offered a fellowship in the first four rounds of FLF.

an interview in each round. Even using sift scores, rather than interview scores, to identify those at the margins of treatment, with only 550 fellows, it is possible that the numbers scoring around the treatment threshold will be insufficient to detect any impact from FLF. For this reason, a RDD would only be recommended to supplement the main PSM design and to assess the likely robustness of the evidence on the impact of FLF.

#### **Box 4: Other methods of identifying causal impact**

This box provides brief details of a number of other methods which are commonly used to identify the causal impact of policy interventions and explains why these approaches are not suited to estimating the impact of FLF. See also Appendix C.

##### **Randomised control trial (RCT)**

It would be necessary to identify applicants who met the criteria to be offered a fellowship and then assign them to the treatment and control groups at random. With random assignment to FLF, outcomes for the control group would be likely to provide a robust estimate of the outcomes that fellows would have attained if they had not been treated. This is because there would be no reason to expect any systematic difference between the two groups which would affect outcomes following random assignment. However, a RCT is not feasible for FLF due to the fact that most fellowships have already been awarded and so the numbers of applicants who could be randomised is likely to be insufficient to detect any impact from FLF.

##### **Difference-in-differences (DiD)**

With this approach, the trend in outcomes for fellows prior to the award of the fellowship would be compared with the trend in outcomes for a comparison group. The assumption is that both groups would continue to experience a similar trend in outcomes following the fellowship, and so any divergence in trends that emerges after the treatment group receive the fellowship can be attributed to the impact of FLF.

DiD analysis is most likely to provide a credible estimate of impact where there are restrictions on eligibility for an intervention which create a pool of potential comparators who are likely to be similar to the treatment group and to experience a similar trajectory in outcomes over time, but are prevented from taking part. As there are few restrictions on who can apply for FLF and it is a UK-wide scheme, it is difficult to identify a suitable comparison group.

A DiD analysis would also require access to longitudinal data on outcome measures before and after the intervention for both the treatment and comparison groups, including a long-run of pre-intervention data to test whether the assumption of common trends is met. As there is no single suitable data source which covers all FLF participants and a potential comparison group, it would be difficult to meet the data requirements for a DiD analysis.

##### **Survival analysis**

Where the number of applicants meeting the criteria to be offered an intervention exceeds the number of places available in the short-term, it may be necessary for suitable candidates to wait to start an intervention. Outcomes for those who start to receive the treatment at an early point in time can be compared to those for individuals who have to wait to receive the intervention.

The impact of FLF is only likely to emerge over a matter of years and the funding rounds are around six months apart. As a result, it would only be feasible to obtain an estimate of the impact

of FLF using survival analysis over the very short term, or alternatively, if successful applicants could be made to wait a matter of years before starting the fellowship.

### 6.2.3 Approach to data collection

The data collected during the application process for FLF will be used as the primary source of baseline information on the treatment group and the main comparison group of unsuccessful applicants for FLF. This could potentially be supplemented by data collected as part of the application process for other fellowship schemes to carry out an analysis of the relative effectiveness of FLF compared to other types of fellowship.

It is recommended that information on the outcomes experienced by participants in FLF, as well as the comparison group, be collated from secondary data sources (see Chapter 5). Whilst it may be possible to use the contact details provided by applicants to conduct a bespoke survey of the treatment and comparison groups, it is likely to be difficult to maintain contact with the comparison group in particular over time. FLF seeks to support fellows for a matter of years (between four to seven years) and it is apparent from the theory of change that long-term impacts are anticipated. Whilst it may be straightforward to maintain contact with those in receipt of funds from FLF, at least during the period that they are directly benefiting from the fellowship, it is likely to be much more difficult to keep in touch with the comparison group and keep them engaged in primary data collection (they are less likely to respond to survey requests about a programme that they were unsuccessful in gaining a place). As one of the aims of FLF is to increase employment stability, the comparison group may also be more likely to change jobs than those offered fellowships. As a result, it is likely that survey response rates will be lower for the comparison group than for fellows. If those in stable employment are more likely to respond to the survey than those who are unable to find permanent work, the estimated impact of FLF will be biased.

Using secondary data sources to observe outcomes for the treatment and comparison groups reduces the risk that impact estimates will be biased by attrition. It is also likely to be feasible to observe outcomes for a large proportion of applicants over a longer time-frame than would be possible with a survey. As it would not be necessary to ask FLF applicants to complete a survey, it would reduce the burden on applicants and increase the feasibility of producing regular updates to the impact estimates by avoiding the need to conduct further rounds of survey research.

To be able to use secondary data sources to estimate the impact of FLF it is essential to make applicants aware that their personal data will be stored and used to track their outcomes over time when they apply for FLF (as discussed above). If this is not the case, the approach will not be feasible. The ability to use secondary data sources to observe outcomes also depends heavily on whether the information supplied by applicants is adequate to identify their records in other datasets. If the personal data are of poor quality or do not contain sufficient detail to unambiguously identify records for the right individual at the outset, it will not be possible to proceed with the impact evaluation. Even if it is possible to identify records for a large proportion of applicants initially, it may be difficult to track them over time if they change job, location, move out of research or innovation or if other



personal details change. Some attrition is likely over time, even when observing outcomes in secondary data sources and this may make it difficult to obtain a representative estimate of the long-term impact of FLF.

A further disadvantage of using secondary data to observe the impact of FLF is that it is not possible to tailor the outcome measures to ensure that they capture the full range of expected impacts from FLF. Rather the analysis will be limited to estimating the impact of FLF on outcomes which can be observed in pre-existing datasets. As the expected impacts are wide-ranging and applicants will come from business and academia, it will be necessary to collate data from many sources and this is likely to be resource-intensive. Nevertheless, the potential disadvantages of relying on bespoke survey data are likely to outweigh the disadvantages of using existing secondary data and so it is recommended that the impact evaluation focus on compiling outcomes measures which can be observed in secondary data sources.

## 6.2.4 Estimating impacts at different levels

It is apparent from the Theory of Change and the outcome measures discussed in Chapter 5 that FLF is expected to have impacts at many different levels, including on individual fellows, research teams, departments, host institutions, external collaborators, the research and innovation community and wider society in the UK and abroad, as well as the UK economy. Whilst outcome metrics may exist to estimate the impact of FLF at different levels, it is necessary to consider whether in all cases it is credible to attribute any differences in observed outcomes between the treatment and comparison groups which emerge over time to the impact of FLF alone.

For example, if a fellow works for a large multinational corporation, it is unlikely to be justified to regard any difference in the financial performance of the organisation compared with that of organisations employing matched comparators as due to the impact of FLF alone. In this case it is likely that there will be unobserved factors which explain differences in organisation-level outcomes between the treatment and comparison groups. Furthermore, it is likely to be difficult to find well-matched comparators where fellows work for atypical host organisations, such as large firms in the business sector.

A further consideration when assessing the impact of FLF at higher levels of aggregation is whether the scale of support provided is likely to produce an impact that is sufficiently large to be detected at that level. For example, in a very large organisation, any impact from a single employee being awarded FLF would have to be substantial for outcomes to be affected at the organisation-level.

Whilst it will be important to explore the impact of FLF at different levels to ascertain whether it has achieved the wide range of objectives identified in the Theory of Change, to avoid wasting evaluation resources in sources and collating data it would be advisable to pare down the list of potential outcome measures (outlined in Chapter 5) to use in the counterfactual impact evaluation to those where:

- there is a strong case for expecting impacts to be seen at a given level; and



- it is credible to believe that any differences in outcomes between the treatment and comparison groups are due to the impact of FLF.

## 7 Considerations for process evaluation of FLF

This chapter considers the feasibility of robust evaluation through using process and theory-based approaches. The reality is that these methods are far more flexible than counterfactual impact evaluation approaches and therefore there is no real question that they can be applied effectively. The feasibility therefore considers what is most desirable in respect of rounding out the picture on impact to demonstrate the programme's effect, and of these desirable approaches, what adds most value in respect of the insights generated.

### Key findings for an evaluation design for FLF:

- FLF is an ambitious and complex programme with impacts being intermediated separately by fellows, hosts and R&I ideas (which in combination are also an engine for change) with EDIP being a strong lever in all causal pathways.
- The feasibility has established it is possible to mobilise robust methods to address whether the FLF has the planned impacts on the outcomes identified for fellows (which also start to identify outcomes related to hosts and ideas). However the impact method cannot address the causal pathway for the fellows, or the issues that may be encountered in the small business segments in respect of drawing suitable comparisons. Theory-based evaluation therefore becomes the best option to gather data on these issues.
- Process evaluation adds value and context – the latter is supportive in the context-mechanism-outcome analysis undertaken for theories of change. The process evaluation will also surface information about the inputs to the programme and the activities that support it – which will start to unpick how the ambitions for the practice are realised in practice and the implications.
- A final strand of work relates to strategic assessment, which can focus on the unique contribution of FLF to the R&I landscape in the UK. This aspect will also assist understanding the programme mechanisms that had greatest influence in achieving the planned outcomes.

### 7.1 What process and theory-based evaluation add

The counterfactual impact evaluation as set-out in Chapter 6 focuses on the fellow as agent of change, and thereby the causal pathway associated with their journey to outcomes and intermediate impacts. This will however take into account outcomes and intermediate impacts for fellows as well as some that stem from their R&I idea (and from which final impacts will more certainly derive) such as patents and citations. The counterfactual impact evaluation will exploit the applicant group to draw comparison groups for the fellows (ie. those who were successful in gaining FLF funds) using PSM to identify matched cases within the full applicant pool, and RDD to make matches to those fellows

who were shortlisted but did not receive funds following the final stage of the awards process.

Therefore, the process and theory-based research should be designed to complement this by providing detailed insight into the underlying contexts and mechanisms in the fellow's causal pathway that lead to outcomes. It can also supply the main source of insight into the host outcomes and causal pathway (independent of the fellow), and into the operation of equality, diversity, inclusion and porosity within FLF (see Figure 2-2).

### 7.1.1 What are process and critical realist evaluations?

The Magenta book (2020) notes that to be comprehensive, multiple approaches are often required to evaluate interventions. This is particularly the case when interventions are complex and operate over long time periods. A comprehensive evaluation, through using multiple approaches and lenses to understand interventions, can identify whether or not an intervention has an effect over what would have happened anyway, it will also set out why and provide information into how this effect could be replicated. Contributing to this holistic understanding, process evaluations track implementation whereas theory-based approaches are concerned with explaining outcomes and events focusing on how and why these occur (see Chapter 4, and Appendix C for more detail).

### 7.1.2 Objectives for process and theory-based evaluation

The process evaluation of FLF will deepen insights into the delivery of the programme, ensuring key decisions and effects are captured and documented so that the role of these in leading to programme impacts can be fully considered. This will provide explanatory information to the impact evaluation of the fellows' causal pathway as well as provide the key insight into the R&I idea, host and EDIP causal paths and outcomes (see Figure 2-2).

The focus of the process evaluation will be to:

- Understand, describe and document the development of the FLF programme and particularly funding decisions, and the support package offered to fellows, over time
- Gain strategic assessment of the R&I landscape over time, degree of change being seen and likelihood that this can be attributed to FLF
- Gain strategic and operational insight on the role of the programme in generating the outstanding R&I talent of the future
- Gather views on the effectiveness of the programme in delivering the impacts in respect of knowledge, economy and society.

The process evaluation can capture and document how the dynamic development programme changes over time and the reasons for this, providing insights into the different needs of different cohorts as well as the learning that is drawn from support fellows in each funding round. It can also demonstrate the role UKRI plays in creating the circumstances for collaboration and porosity.

The theory-based evaluation will focus on:

- 
- Understanding whether the programme recruits fellows and hosts that link to the ambitions for the programme in respect of range, diversity and porosity (the EDIP pathway and outcomes).
  - Deepening insights by discipline and MIDRI mix, and strengthen evidence for the small segments within the FLF population where counterfactual impact evaluation may not prove possible.
  - Understanding the effect of FLF on fellows' behaviours over time (do they value diversity, collaboration and porosity and positive competition; does this lead to embedding more support and action for diversity, collaboration and porosity? (spillovers). Do fellows intend to continue in the UK R&I context?)
  - Understanding the effect of FLF on hosts' behaviours over time (are hosts retained in UK R&I; do hosts value diversity, collaboration and porosity and positive competition; does this lead to embedding more support and action for diversity, collaboration and porosity?)
  - Understanding whether in combination the evidence indicates that these are leading to acceleration or increased impact of ideas on the economy, or on the retention of high quality international R&I talent (likelihood that longer term impacts will emerge).
  - Deepening insights into what works for whom in what contexts, and which combinations of context and mechanism lead towards outcomes being achieved.

The questions for the theory-based component allow for the causal pathways to be tested in order that they become more specific. For example, leading survey and qualitative research with fellows and hosts can detect whether the programme/funding scheme was attractive because it provides opportunity to think beyond the boundaries of specific disciplines; similarly, qualitative nested case studies focused on fellow (and team), host and their UKRI contact can demonstrate if and how the partnership works as an engine for change in the programme.

The theory-based evaluation can also seek to detect signs of impact (although this will not meet the 'gold standard' of counterfactual, impact approaches). It can use innovative and creative methods that lend weight to narrative and qualitative approaches, such as contribution analysis and process tracing, strengthening the inferences that can be drawn. When the evaluation is commissioned it should be expected that organisations bidding to lead the work would propose the approaches to analysis that they believe would best address theory-based impact. There is also the possibility to draw comparison with peer researchers and innovators through a matched case methodology drawing on sources of administrative data. This latter approach is discussed below, alongside a key method to understand the causal pathways within FLF.

## 7.2 Analytic framework for the process evaluation

**Table 7-1: Analytic framework for the process and theory-based evaluation evaluation**

Research question	Method	Issues (& links to the TOC)
<b>To what extent (and how) is the FLF programme working and being delivered as intended?</b>		
(To what extent) are targets for FLF inputs and outputs being met?	<ul style="list-style-type: none"> <li>• Management information (MI) analysis</li> <li>• Stakeholder interviews</li> </ul>	<p>Fund value, and level of funds distributed to fellows in each of the established segments</p> <p>No. of applicants and fellows awarded, cohort make-up e.g. international and UK, location within R&amp;I sector, in each round and overall</p> <p>Sustained support and commitment from government and sector</p> <p><i>(inputs and early outputs on EDIP ie. diversity of fellows and increased range of hosts &amp; in respect of the difference made by sustained funding creating better terms and conditions for the fellow pathway because it creates certainty in the host pathway)</i></p>
(To what extent) has FLF stimulated multi- and interdisciplinary research and innovation (MIDRI) fellowships?	<ul style="list-style-type: none"> <li>• MI analysis</li> <li>• Stakeholder interviews</li> <li>• Case study exemplars illustrating partnership and 'engine for change'</li> <li>• Peer matched case method</li> </ul>	<p>Characteristics of fellows, hosts and projects in each round and overall</p> <p><i>(additive effect in respect of diversity of fellows/PIs EDIP pathway)</i></p>
(To what extent) are FLF fellowships additional to other schemes supported by UKRI and other UK / international funders?	<ul style="list-style-type: none"> <li>• Stakeholder interviews</li> <li>• Comparative analysis against existing evidence</li> <li>• Career tracker survey</li> </ul>	<p>Views and evidence on uniqueness of FLF contribution</p> <p>Common measures in career tracker surveys to benchmark to other programmes/funding schemes</p> <p><i>(sustained funding creates conditions for fellows to thrive; progress ethos and dynamic development programme support attitudinal and behavioural changes in fellows with positive effects on EDIP; certainty for hosts mean they sustain a role in R&amp;I; hosts see the value of greater porosity and collaboration)</i></p>
(To what extent) has FLF delivered effective post award management to support the professional development of the fellows?	<ul style="list-style-type: none"> <li>• Programme documentation</li> <li>• Career tracker survey</li> <li>• Qualitative interviews with fellows</li> </ul>	<p>Content of the programme over time</p> <p>Views on the composition and effectiveness of the dynamic support programme</p> <p><i>(links to activity by UKRI to embed activities aligned to its mission that create the conditions for researchers, hosts and R&amp;I to flourish; and creates a context for increased porosity and collaboration as well as diversity amongst PIs)</i></p>

Research question	Method	Issues (& links to the TOC)
<b>To what extent (and how) is the FLF programme working and being delivered as intended?</b>		
	<ul style="list-style-type: none"> <li>Stakeholder interviews (operational)</li> <li>Peer matched case method</li> </ul>	
What, in practice, is felt to be working more / less well regarding the delivery of the fund, and why?	<ul style="list-style-type: none"> <li>Stakeholder interviews (operational and strategic)</li> <li>Qualitative interviews with fellows</li> <li>Qualitative interviews with hosts</li> </ul>	<p>Views on delivery – what is working well, what is less effective.</p> <p><i>(linked to outcomes including whether FLF has created multiple cohorts of outstanding researchers who are retained in UK R&amp;I – Fellow's pathway; whether there is increased focus/valuing of diversity within the R&amp;I sector - EDIP; whether increased networking and porosity has emerged (EDIP); and whether there is evidence of increased collaboration in the sector - hosts)</i></p>
What are the unexpected barriers or facilitators to the FLF processes and the delivery of the anticipated outcomes, if any?	<ul style="list-style-type: none"> <li>Stakeholder interviews (operational and strategic)</li> <li>Qualitative interviews with fellows</li> <li>Qualitative interviews with hosts</li> </ul>	<p>Views on delivery – what in the context is supporting/preventing the achievement of outcomes.</p> <p><i>(assessment of outcomes and whether there any are unexpected or not anticipated in the TOC by any pathways and potential effect on intermediate and final impacts)</i></p>
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?	<ul style="list-style-type: none"> <li>MI analysis</li> <li>Stakeholder interviews</li> <li>Career tracker survey outcomes</li> </ul>	<p>Descriptive analysis of who (hosts and fellows) reached by FLF.</p> <p>Views and interpretation on programme targeting and implications.</p> <p><i>(inputs related to EDIP – diversity of fellows/PIs and of hosts; how R&amp;I ideas relate to key strategic themes in industrial strategy; potential for other social and economic outcomes to emerge from the R&amp;I)</i></p>
What lessons are there for future rounds / similar schemes?	<ul style="list-style-type: none"> <li>Stakeholder interviews</li> <li>Meta/comparative analysis</li> </ul>	<p>Benchmarking to other schemes.</p> <p>Views on lessons gleaned by key stakeholders.</p> <p><i>(how the funding round decisions contributed to the outputs in respect of fellows, EDIP, hosts and R&amp;I; how the dynamic development programme and its evolution over time increase networking and collaboration to support the programme to achieve its planned impacts through the causal pathways it embedded)</i></p>

## 7.3 Design

The process and theory-based evaluation must focus on generating insights into inputs, activities, outputs and outcomes efficiently and capture perceived impacts, maximising existing data where possible. The following components are recommended:

**Table 7-2: Elements of the process evaluation**

Component	Coverage	Lens	When and frequency
Management information analysis	Descriptive analysis of: applicants, and those receiving funding, hosts, discipline(s)	<b>Inputs</b> Diversity of applicants and fellows Disciplines and porosity from application to funded fellowship Traditional/non-traditional hosts	<ul style="list-style-type: none"> <li>• Following each funding round</li> <li>• Updated regularly and at minimum, contract breakpoint and end point (to understand length of engagement)</li> </ul>
Stakeholder and operational interviews	Strategic insight into programme delivery and effectiveness with interview questions tailored to the point of delivery Stakeholder sample tailored to be appropriate to the temporal order implied by research questions (see table 5.1); should cover stakeholders: <ol style="list-style-type: none"> <li>a) Involved in decision-making on awards,</li> <li>b) Involved in operational delivery/support;</li> <li>c) At a strategic level;</li> <li>d) Wider stakeholders for comparative analysis; and,</li> <li>e) National/ policy stakeholders (eg BEIS)</li> </ol>	<b>Inputs and activities</b> Factors feeding into funding round decisions Support and networking for fellows and hosts <b>Outputs/ outcomes/ impacts</b> Reflections on achievements of FLF; plausibility of causal mechanisms <b>Context</b> Factors affecting decisions, plans and outcomes at each phase of interview	<ul style="list-style-type: none"> <li>• Following each funding round (6 months)</li> <li>• Moving to a schedule aligned critical points – contract extension point for each cohort; end of funded period for each cohort.</li> <li>• Aligned to intermediate and final impact points</li> </ul>
Fellows – career tracker survey	‘in programme’ experiences, views, attitudes, and behaviours Longer term career trajectories	<b>Inputs and activities</b> Starting points including ambitions for self, host and R&I idea; and attitudes Experiences of support from UKRI, hosts and networking <b>Outputs/ outcomes</b> Effect of support, funding, networking	<ul style="list-style-type: none"> <li>• Annual survey ideally</li> <li>• Contract extension point for each cohort;</li> <li>• End of funded period for each cohort.</li> <li>• Longer term tracking of career</li> </ul>



Component	Coverage	Lens	When and frequency
		<p>Changes in attitudes and behaviours</p> <p>Difference made by FLF</p> <p><b>Longer terms effects on careers</b></p> <p>Tracking career transitions and progress over time, ideally using common benchmarks with other fellowship schemes</p>	
Hosts survey	<p>'in programme' experience, views, attitudes, and behaviours</p> <p>Aimed at sponsor/supervisor (senior strategic manager)</p>	<p><b>Inputs and activities</b></p> <p>Starting points including ambitions for fellow, organisation and R&amp;I idea; and attitudes</p> <p>Experiences of supporting fellow including use of networks and links to UKRI</p> <p><b>Outputs/ outcomes</b></p> <p>Effect of support, funding, networking</p> <p>Changes in attitudes and behaviours; future intentions</p> <p>Difference made by FLF</p> <p><b>Longer term effects</b></p> <p>Intentions and behaviour on ongoing engagement with R&amp;I</p>	<ul style="list-style-type: none"> <li>• Contract extension point for each cohort;</li> <li>• End of funded period for each cohort.</li> </ul>
Fellow and Host interviews and nested case studies	<p>Examples of the interaction between fellow and host, the effect for the idea, and role of UKRI ie. the 'engine for change'</p> <p>Sample drawn to demonstrate the segmentation; and different dimensions of porosity (MIDRI)</p> <p>Case studies can be thematic focused on porosity, collaboration, as well as showcasing the partnership and difference made by being part of the engine for change</p>	<p><b>Outputs/ outcomes</b></p> <p>Effect of support, funding, networking</p> <p>Changes in attitudes and behaviours</p> <p>Difference made by FLF</p>	<ul style="list-style-type: none"> <li>• Towards end of funding period so that ideas are coming to fruition and interviews are reflective (not prospective)</li> </ul>
Fellows and peer matched-cases	<p>Using alt-metrics to provide insights into the similarities and differences between the careers and outputs of fellows and matched peer researchers/innovators.</p>	<p><b>Outcomes and effects</b></p> <p>Differences/similarities in outputs (patents, papers, etc)</p> <p>Differences in behaviours (linked to MIDRI porosity)</p> <p>Difference made by FLF</p>	<ul style="list-style-type: none"> <li>• Mid-term and towards the end of funding period so there is more chance that intermediate and final outcomes can be observed</li> </ul>

## 7.4 Optimising the design

As noted earlier, all the methods noted above are feasible and practical to implement. Each will contribute to the broad understanding required from the process and theory-based evaluation of how the programme operated and caused (or did not cause) the outcomes and impacts that result. Evaluation designers can reach decisions on theory-based analytic approaches that can best capture the causal pathway, with the options presented in the appendices drawn from the Magenta Book are all able to add value.

It is recommended that the process and theory-based evaluation documents and tracks delivery of the programme, 'describes' the cohorts of fellows and hosts, tracks the inputs and activities they receive and their movement through outputs, outcomes and towards intermediate and final impacts are captured. As there is a limited evidence base for the causal pathways identified in the theory of change – because evaluation of precursor fellowships is limited, and because of the unique intentions for and thus features of FLF – the process and theory-based evaluation provides the best means to test these and understand how the programme works in practice. This includes particularly the softer outcomes for fellows related to leadership and being inclusive that would not emerge from counterfactual impact evaluation; as well as detailed evidence on the outcomes for hosts and how well the theory of change predicts these; and similarly related to EDI/porosity and crucially the idea itself.

Decisions therefore about what to include in the process and theory-based evaluation design concern what will generate the best insights on the budget that can be made available. The qualitative research is crucial as without these accounts of practice, and reflection on delivery, key lessons for the future could not be drawn. Operational and strategic assessments of programme delivery will reveal how well the inputs and activities were matched to the ambitions in the TOC and whether these then had a causal relation with the planned outputs and outcomes, which would increase the plausibility that intermediate and final outcomes might be attributable to the programme.

Practice examples (case studies) would enrich programme data with illustrations of how the partnership engine plays a role in producing outputs related to the R&I idea as well as EDI pathway. These are relatively simple to generate. Importantly, consents are in place that enable contact with the fellows themselves in primary, qualitative research and it is likely that stakeholders will wish to engage to share views. While there is not a clear consent model for hosts, it is likely they could be engaged through the fellows involved in qualitative research, providing a nice triangulation between perspectives on outcomes in case studies.

In contrast, the fellow career tracker survey and the host survey would require a higher level of funding, with costs increasing with the frequency of repeat surveys as analysis becomes more complex. There are a few considerations in whether and why to take forward surveys and how frequently, discussed further in section 7.5. However, both are required to fully understand the range of outcomes for fellows, and specifically to track whether the programme – and its funding – creates the certainty for hosts to be retained, to

offer better contractual terms to help retain R&I talent and value and thereby increase collaborative working.

A novel approach, that sits within the theoretical suite, is to use a matched-case method to understand and record the outputs of FLFs compared to a selected peer researcher to understand if there is an increased propensity to work in new collaborations with a MIDRI focus ongoing, and to understand more of the effect on porosity in the system. Similarly this approach could bolster evidence for fellows in the ‘business’ segmentations where impact evidence is likely to be weaker. More detail on this approach is supplied in section 7.6. below.

## 7.5 Tracking surveys

### 7.5.1 Tracking fellows

A first point to make is that as the fellow survey would track only successful applicants (not unsuccessful ones) since, in the counterfactual impact evaluation, evidence of outcomes will draw on administrative data which will increase the likelihood of identifying outcomes for a large proportion of the cohort, whereas a survey would be affected response rates.

The survey in the process evaluation encounters few of the practical disadvantages that it faces for a counterfactual impact analysis using the RDD approach. In the impact evaluation, prospects for tracking unsuccessful applicants relative to the fellows are much reduced, thereby affecting whether robust estimates of impact could be drawn. In the process evaluation, the experiences of the fellows is the focus of interest in order to capture evidence on attitudes and behaviours as well as softer outcomes that will not be evidenced in the administrative data.

Fellows are likely to be highly engaged – they have received long-term funding and support from UKRI and the FLF dynamic development programme to develop their careers and R&I, which is likely to act as an incentive to survey completion. Establishing an annual survey will provide for deep insights both into programme experiences and outcomes and will be informative in respect of time-lines to outcomes in different discipline mixes. While there will be some attrition between survey waves, a longitudinal panel offers some statistical advantages in analysis; and evidence from the evaluation of other schemes, indicates that on some measures, attrition can be overcome through manually tracking non-respondents in administrative data-sets. Moreover, in the process and theory-based evaluation, if career tracking survey of fellows were to use a common framework with other career tracker surveys, this provides an option for comparative analysis of outcomes (effects of FLF compared to other similar schemes).

### Overview of and content of tracking surveys in other evaluations

An approach used in previous evaluations of funded research Fellowships or Chairs is to track the ongoing careers of funding recipients via a survey or series of surveys that take place after the period of funded research has been completed (see Chapter 4 and Appendix D). The funding recipient may be asked to provide information about their current

role and employment history, their research achievements, their experience of the funding programme and other goals of the programme such as outreach, collaboration or dissemination. Several programme evaluations have collected information in areas of particular relevance to the aims of the FLF which suggests opportunity to draw on their question sets to generate common measures:

- In a recent survey evaluating a range of early career fellowships, senior fellowships and chair funding programmes, the *Royal Academy of Engineering* explored collaborations and long-term relationships with industry including understanding of the partner organisation's culture, activities engaged in with the partner since completion of the funded-programme, eg. subsequent joint research, setting up a company, and plans/intentions for further engagement with the partner organisation or the sector.
- Similarly, the *Sir Henry Dale fellowship* funded by the Wellcome Trust and the Royal Society aims to support early career researchers in biomedical research to become independent research leaders and its evaluation survey explored activities and outputs evidencing leadership at wider societal level such as contributing to policy-making or public engagement, eg. public lectures.
- The *Newton International Fellowship (NIF)* programme, funded by the Royal Society, British Academy (BA) and Academy of Medical Sciences (AMS), aims to attract the best early career post-doctoral researchers world-wide to conduct research in the UK, as well as fostering long-term collaborations with researchers in other countries and build research capacity. A recent evaluation survey included questions which explored continued engagement with their UK host institution and research activities in the UK, attitudes to international collaboration and intentions/expectations around further engagement with the UK.
- The *Principal Investigators and Research Leaders Survey (PIRLS)* managed by Vitae identifies a range of activities that research leaders undertake including academic collaborations (including interdisciplinary and international), collaborations outside HE, public engagement and outreach activities, developing researchers, and providing career development advice to others on careers outside higher education.
- The *Royal Society's recent inaugural Career Tracker survey* compared aggregate results from the PIRLS survey with relevant aggregated results from their own tracker.

While the scope and objectives of the FLF programme are quite broad and far-reaching compared with other fellowship programmes, other programmes also focus on areas and objectives of relevance to the FLF and can provide examples of information capture around these topics. Topics have included:

- **Career destinations:** current employment and career destinations, next steps after completion of the fellowship, eg. first role after a fellowship. For the FLF, this would help to evidence whether fellows are sustained in R&I, whether their tenure is of longer duration and whether that allows them to rise to become the inclusive influential people the TOC envisages.

- **Perceived impacts:** on their career, perceived benefits of the fellowship, perceived difference to their development as a researcher. This again would capture perceived effects of FLF (and additionality) which could be compared to other programmes.
- **Research activities:** Research outputs/activities, grants or funding awarded since the fellowship, awards or prizes received, outreach/leadership, research outputs for their research group/lab. Gathering this data in the survey would enable triangulation with the administrative data as well as some of the softer leadership outcomes to be observed.
- **Collaborations:** ongoing collaboration with the host organisation, collaborations in the UK (if international), international collaborations, collaborations with industry. Capturing this for FLF would add to the evidence base on EDI and porosity.
- **The Fellowship programme:** reasons for applying to the fellowship programme, experiences of the application process, other programmes applied to, experience of the fellowship, training/mentoring opportunities. For the FLF, this information would generate insights into the prestige of the FLF, which is an input for the TOC.
- **Career motivations and intentions:** plans and intentions of those nearing/at the end of the fellowship, intentions of individual at their current career stage. For the FLF, this will help identify effects in terms of retaining fellows in the R&I sector.

In the process and theory-based evaluation, the focus is on career tracking, and perceptions of impact. The surveys will capture outcomes measures that are less tangible and less well covered by the hard metrics to ensure the fellow causal pathway is fully elaborated. As noted, we are not proposing this because of the likelihood that unsuccessful applicants will be less likely to engage in surveys.

### Approaches to survey timing in other evaluations

There has been wide variation in the timing of career tracker surveys. Some have been one-off snapshots meaning that some cohorts were in the process of undertaking the fellowship and other cohorts had completed their period of funding a decade or more earlier, eg. Royal Academy of Engineering evaluation. This approach could be emulated in the FLF, although all cohorts remain in their funding period and data would not be informative on longer term outcomes.

Others, such as the Wellcome Trust are implemented as an annual approach. For example, the Wellcome Trust runs career trackers for three main groups of funded researchers: the Basic Science Career Tracker (BSCT) which includes PhD students, postdoctoral sciences and researchers returning after a career break, the Clinical Career Tracker which includes PhD students, postgraduate training, postdoctoral and senior researchers in clinical fields, and the International Career Tracker which includes interns, Masters and PhD students, postdoctoral researchers and senior researchers in low- and middle-income countries (LMIC). All follow a similar paradigm with the BSCT running the longest, from 2009 with the 2003-04 PhD cohort. Grantholders are invited to complete the survey in the final year of their grant and the Wellcome Trust plans to track cohorts for at least ten years. In this way, the Tracker can build up a rich picture of career intentions and development over a ten year period to capture immediate and longer term impacts of the funding. For the FLF, the particular benefits of this would be adding specificity to the causal

chains and timelines in respect of outcomes emerging for different disciplines and MIDRI project mixes, as well as for the four segments in the population we have identified.

The Royal Society is using an intermediate approach for its Career Pathway Tracker which follows recipients of their University Research Fellowship and Dorothy Hodgkin Fellowship. These fellowships have been running for 38 and 26 years respectively. After undertaking a survey of all previous grantholders in 2017/18, the Society intends to survey recipients every five years to track careers over the longer term. While this approach could be emulated within FLF, less insight would emerge on the relationship between programme activities and outputs, outcomes and impacts and thus less evidence for the causal chains identified in its theory of change.

### **Risks of attrition and approaches to improving response rates**

UKRI already conducts some surveys with current Future Leaders Fellows and current recipients of funding may be highly motivated to complete surveys for their funders but, as noted earlier, there is a risk of attrition with longitudinal studies and the relationship of a former funding recipient with their funder may weaken over time. The most recent publicly available data for the Wellcome Basic Science Career Tracker (BSCT) details findings from Wave 6 in 2014 and provides an overview of response rates for the different funding programmes over this time. Participants from different funding programmes joined the tracker at different points; the earliest cohorts invited to take part were those who started an International Senior Research Fellowship in the 2001/02 academic year and the most recent cohorts started in 2009/10 on either a PhD programme, Sir Henry Wellcome Postdoctoral Fellowship or a Research Career Development Fellowship. The Wellcome Trust report that the BSCT has achieved an annual response rate of around 80 per cent since it started in 2009.

A key strategy the Wellcome Trust has used to increase response rates is to supplement the online survey by web research to find individuals who do not respond. For the 2014 BSCT, this increased the response rate from 78 per cent with survey responses to 89 per cent, which increased the number of respondents from 713 to 811 out of total of 912 grantholders that were invited to take part. The Wellcome Trust also offers an incentive to those who have been included in the BSCT for at least five years (since wave 1 or wave 2) of either a charity donation or a £10 Amazon voucher for completing the survey.

We must note that some evaluations have included tracker surveys as part of work to establish a comparison group. Some involve a comparison group in their career tracker surveys by surveying (all) unsuccessful applicants as the comparators in order to illustrate the gap that can emerge between those receiving funding and those who do not. However the response rates for unsuccessful applicants can be very low. One example is the evaluation of the Springboard Award where the response rate for successful applicants was 72 per cent, compared to just 27 per cent for unsuccessful applicants.

### **Pitfalls in longitudinal surveys**

Despite these encouraging signs, there are some disadvantages or pitfalls to avoid when implementing longitudinal tracker surveys. One is that there is likely to be an element of

self-selection in those who answer the survey. For example, those who have gone on to experience relative success with their R&I may be more motivated to respond and maintain a relationship with their funder, whereas those whose progress has been more modest or who have moved away from R&I may be less inclined to respond. Similarly, using web research to find individuals and their information may be a more successful strategy for those in academic roles or with more traditional research outputs, eg. publications, than for individuals who are working in industry or who have moved away from research entirely. Using administrative data to understand outcomes in the counterfactual impact evaluation will counteract at least some these issues.

Furthermore, research activities and development taking place in the context of industry may not be publicised in the interests of competitive advantage, or may not be visible externally as they feed directly into further development of products or services rather than being disseminated through papers and conferences. Hence, even in the process evaluation and in assessing perceptions of impact and effects of FLF there will be some weaknesses in using a survey approach in the evaluation. To counter this or at least understand the likely skew that is emerging, it would be important to lead an analysis of non-responders to identify any common, observable characteristics among those who do not respond but there remains a risk that non-observable characteristics drive non-response.

Another possible challenge for a longitudinal survey would be the potential high mobility amongst the fellows and particularly, maintaining responses from fellows who move away from the UK over the course of their career. Data trees produced for Wave 5 of the Wellcome Trust's BSCT include a number of individuals who moved away from the UK (and in some cases back) which indicates that it is possible to track grantholders through international moves but it is unknown what proportion of those who did not respond to the survey had moved internationally and whether this played any role in non-responses. Positively, the response rate from the International Senior Research Fellows subgroup in the BSCT is at a similar level to other subgroups (82-100 per cent) across the year cohorts and the response rate for the 2014 International Career Tracker was 100 per cent although this is a very small group (N=10) which suggests that working outside of the UK does not appear to a significant barrier to responding to the survey.

Finally, it should be noted that tracker surveys rely on self-report and while this does not necessarily mean that information submitted is necessarily inaccurate, other data may be needed to explore, validate and/or understand activities or outputs reported, e.g. the extent or financial value of collaboration with industry.

### **Recommendations for a tracker survey in FLF**

Using a career tracker survey as part of the FLF evaluation would allow UKRI to explore some of the diversity of destinations open to Fellows as well as understand more about the softer outcomes eg. on leadership, diversity and porosity. As the programme aims to enhance collaborations and movement between academia and industry, as well as multi- and inter-disciplinary ways of working, career tracking would capture some of this in practice. A data tree produced for the Wellcome Trust's Basic Science Career Tracker



shows how some individuals move back and forth between academia and industry over the course of five Waves (years) of the tracker.

Taking an annual approach to tracking would enable a rich picture of Fellow's activities during and after completing the fellowship but it is the most resource intensive option. This would include responses to the development programme, as well as tracking attitudes and behaviours concerned with EDI, porosity and collaboration as well as leadership outcomes.

However, as the FLF programme, covers a broad range of research areas including Arts and Humanities, Social Sciences and STEM, and crucially multi- and inter-disciplinary R&I, it is difficult to identify the optimal interval at which to carry out the tracker as the typical length of time required to achieve research outputs varies widely by discipline, not least due to outputs themselves varying by discipline. For example, a 2019 report from Research England identified that the format of research output differed by discipline with relatively quick forms of dissemination such as blogs, social media and podcasts more common with researchers in the social sciences and arts and humanities, whereas more researchers in medicine, health and life sciences, and physical sciences, engineering and mathematics produced peer review – which necessarily take a far longer time to be published.

Annual data collection would allow UKRI to observe and identify any such patterns across disciplines, MIDRI mixes, and types of output for the FLF objectives. As the focus of the FLF centres on leadership and innovation, it may take longer for such outputs to establish themselves but ongoing tracking would help identify any intermediate outputs that build up to bigger changes in sectors, culture or wider society.

Longitudinal surveys offer some statistical benefits for understanding and measuring change, because there is potential to analyse for change at the individual level and in aggregate in the same sample. At the individual level, longitudinal datasets enable statistically significant levels of change across waves to be identified from smaller sample sizes than would be possible in cross-sectional studies – ie. longitudinal datasets are more sensitive to change. They can also allow for a detailed understanding of the nature of change. For example, if it was found that satisfaction with the FLF development programme is constant in the aggregate samples, a longitudinal approach can identify if this reflects consistency of experience, or high but equal levels of positive and negative change among different individuals (eg. different disciplines), which could not be directly measured with cross-sectional samples, or purely aggregate comparisons between panel waves. Finally, longitudinal studies can help understand reasons and drivers of change.

A disadvantage of longitudinal research, as noted earlier, is that representativeness is likely to diminish over time due to attrition. The strategy of collating publicly available information on non-respondents can overcome this risk; in addition, the likelihood of high response rates amongst recipients based on other careers trackers suggests the risks of conditioning are minimal.

## Summary

Within the process and theory-based evaluation, a regular tracking survey of fellows would enrich insights from the evaluation allowing tracking through the causal pathway particularly to understand the effects on softer outcomes including valuing MIDRI ways of working, being an inclusive leader and understanding the timelines in the causal pathways

for different disciplines and mixes of disciplines. There is limited foregoing evidence covering these themes – given these are unique features of FLF, and the counterfactual impact evaluation will not be able to detect them (or at least all of them) in the administrative data. Tracker surveys will provide a good insight because previous evaluations suggest response rates are high in similar schemes and fellows will be motivated to report back on the difference the funding is making to their careers.

There are some risks concerned with non-response bias however these can be minimised through desk research to track non-respondents on outcome measures where there is publicly available data. This non-response bias can be a particular cause for concern in longitudinal research with particular groups disproportionately dropped out of surveys, however, the desk research to infill the gaps along with the advantages of being able to analyse change and correlations at individual and aggregate level indicate that in the process evaluation the approach would add value.

### 7.5.2 Tracking hosts

A further consideration for primary research and particularly a tracking survey is the host organisation of Fellows. FLF host organisations can be academic institutions, businesses, or other independent research organisations (as long as they are UK-based), and they play a key role in the successful development of the fellow and the research. The host is part of the ‘partnership engine for change’ which brings together the fellow, host, R&I idea and UKRI (see Theory of Change, Chapter 2). The hosts are responsible for managing the fund, and they are required to commit significant institutional support including: allowing (in most cases) fellows to spend 100 per cent of their working time on the fellowship; a financial contribution to the cost of the Fellowship (20 per cent rising to 80 per cent in academic hosts); and, for fellows in academia, this also includes a commitment to award a UK-based permanent research or innovation position at the end of the Fellowship. However, host organisations can benefit significantly. For example, the anticipated benefits for businesses (outlined in FLF programme supporting documentation and the Theory of Change, Chapter 2) include helping a company to develop and embed new sector or technology expertise or new business model, taking the business in a new direction or opening up new markets through research beyond the normal activity of the business with a ‘higher than normal’ risk.

Given their role in the ambitions for the FLF programme, host organisations are part of the theory of change featuring in the engine for change (along with fellows, the R&I idea and UKRI) as well as having their own causal pathway to support the final impacts. This concerns them being retained in the R&I sector (commitment), valuing porosity and collaboration and these factors feeding through into how they interact in the system as well as build more certainty for PIs. The hosts are part of a complex, overlapping set of causal pathways that lead towards the final impacts on society and knowledge and their outcomes need to be understood at the organisational level and in the context of the engine for change.

A survey of hosts could provide additional quantitative and qualitative data to triangulate and validate fellow’s feedback on outcomes and impacts, and provide additional team and

organisation level data that fellows might not have access to, thereby speaking to another key set of causal pathways and intended outcomes in the theory of change. It could track host outcomes which include: host intentions to continue in the R&I sector, intentions and behaviours around supporting Principal Investigators (PIs), collaboration across organisations and disciplines (which is an intermediate impact for hosts, linked to the EDIP causal pathway and crucially how this leads towards the major impacts expected from the programme which are achieved through combinations of the host and the idea causal pathways). A host survey also could provide evidence into host organisations' motivations, further involvement in FLF (any additional applications or awards), involvement in other R&I funding, involvement in other fellowship programmes and perception of the opportunity provided specifically (and additionally) by FLF. This would contribute to unpicking if and how FLF has enabled new and/or innovative forms of organisational engagement in large, high-quality R&I endeavours.

To generate the detailed information indicated above and by the theory of change, it would be crucial to target an individual with strategic oversight to respond to the survey. This could be aimed at the Fellow's supervisor or the initial sponsor (potentially the same individual). All host organisations are required to provide a Head of Department supporting statement; for academic institutions this is the Pro-VC and for businesses this is the Senior Business Representative.

These surveys could take place less frequently than the proposed annual tracking of Fellows to reduce the burden on the sector (in line with wider moves). A survey at the outset will be important to capture initial motivations and expectations (for impact, causal mechanisms and likely wider influences on impact), and a survey at the end of the initial funding period (4 years) and extended period (additional 3 years) would capture outcomes and impacts and wider contextual influences and record whether the engine for change powered the causal pathways through hosts, fellows and the R&I to final impacts.

It is interesting to note how work to evaluate the European Research Council's programme of supporting early career researchers (including through 'Starting Grants') concluded a key lesson was to be less concerned with evaluating improvements in researcher organisations' performance or strategic capabilities: due to limited capacities and the lengthy period required for resource allocation changes to take effect in organisations; and that impacts are likely to be influenced by conditions and factors beyond the funder's control, so all the funder can do is to optimise the conditions for impact to occur. Therefore, a survey of hosts would also focus on measuring existence of conditions for impact to occur – such as whether the FLF did build stable conditions for R&I, giving certainty and thereby building confidence in those taking forward ambitious programmes of R&I.

Surveys of hosts appear rare and may be limited to monitoring and evaluation work for innovation programmes such as Accelerator Schemes and focused on business sector hosts. For example, the evaluation of DigitalHealth London's Accelerator involved an impact-focused survey with SME beneficiaries (with a relatively low response rate/opt in) and an economic impact survey (undertaken by the delivery agent). The impact survey was closely related to the ambitions for the programme and covered: professionalism of Accelerator staff; whether expectations for the programme were met; perceived reductions in barriers to innovation; perceived closeness to the market (pre and post programme); and

improvements made as a result of the support to date and expectations for the future on aspects such as internal expansion, job creation, R&D spend, further private and public investment, products and services developed (new sales, new customers), new contacts, enhanced cooperation, and additional company turnover resulting from the programme. The survey asked respondents about additionality/deadweight – whether the commercial benefits made resulting from the programme would have occurred anyway with answer categories of not at all, at a later date, by a smaller amount, later and a smaller amount, or the same). The economic aspects captured by survey included specifics attributed to the programme in terms of additional: company growth, employees hired, contracts gained, investment raised, export contacts, savings achieved and beneficiaries. Incorporating perceived impacts of R&I on the economy and society as well as the perceptions of the economic effects would be valuable the FLF host survey, since outcome data available from administrative sources will be limited on these measures.

Our review of fellowship (or similar) programmes and evaluations (see Appendix D) finds that evaluations conducted have sought to include feedback from hosts but they tended to do so through qualitative case studies and interviews. These are useful in capturing detailed and comprehensive information as well as examples, and are recommended for the FLF evaluation. For example, the evaluation work undertaken for the European Research Council of their programmes (as noted above) used case studies to measure the impacts on hosts in terms of: performance – internal funding, organisational decisions, career opportunities; and organisational capabilities – investment in talent, and speed and scope of any changes. These will of course be affected by the starting position of the host and wider context which can make attribution difficult. Another important example where detailed case studies of hosts capture impact is the Research Excellence Framework (REF) exercise – in this case it is academic organisations as recipients of/hosts for funding. A study of REF submissions by KCL (King's College London and Digital Science, 2015) highlight the importance of case studies as they capture the huge diversity of connections and pathways to impact between research and society (which is difficult to reduce to numbers) and illustrate and thus accommodate how different types of HEIs specialise in different types of impact.

There will be trade-offs if only one approach is commissioned – either a survey or qualitative case studies. Surveys can generate systematic information on a range of factors but cannot supply the detailed explanatory content of qualitative research. Given that hosts are part of an engine for change in FLF and have their own causal pathway, it reduces risks of a lack of evidence meaning the evaluation cannot reach judgement if both approaches are commissioned.

## 7.6 Matched case studies

The 'matched pairs' approach, which is located in the theory-based evaluation, would draw on data held in academic and research engines to *approximate* the detailed statistical process of Propensity Score Matching (PSM) in order to identify a relevant comparator for an FLF award recipient (outside of the FLF programme ie not confined to FLF applicants). This approach would use full career history to identify the fellows' peers at the specific point

in time the baseline is drawn. After this the careers of fellows and their matched peers can be tracked within the data held in databases and other online sources. This approach was used in an evaluation of European Research Council programme and 19 matched pairs were identified for further analysis (Nedeva, 2012).

Whilst this approach has the potential to generate a very substantial degree of quantifiable statistical information (citations, promotions, research grants etc), mining through the layers shows it can also identify huge amounts of softer career history and esteem information such as personal awards, advisory roles etc. These softer outcomes – which would not be evidenced otherwise than through qualitative research - map to outcomes identified in the theory of change related to being highly skilled and influence international research leaders, and inclusive leaders who value diversity and MIDRI collaborations.

Moreover, using this approach may help overcome some of the risks that may be faced in the counterfactual impact evaluation of fellows' outcomes, that particularly concern those fellows being hosted by business. The small numbers in the applicant and recipient pools may undermine the evaluator's ability to draw suitable matched cases in either the PSM or RDD approaches. Using the matched peer approach here overcomes that by drawing comparison with other similar researchers and from a wider pool. While, it has to be acknowledged this approach is not as robust as the impact method recommended in Chapter 6, nonetheless, being able to benchmark, using quantitative methods, to a reasonably well-matched comparison group will offer advantages over not being able to draw suitable comparisons at all.

Furthermore, there may be potential for additional quantitative surveys and/or qualitative interviews with the matched pairs about career evolution and development and key factors that shaped the pathway and choices that an individual chose at critical points – subject to personal data being in the public domain in these sources.

### Operationalising the matched pair approach

The approach we recommend would be to draw a matched pair for each of the four segments for the six cohorts. This would lead to 24 matched pairs, and the careers of 48 individual researchers being tracked in the available data.

The information that enables a comparative analysis is provided by the research outcome databases, including general databases covering multiple disciplines and also more specialist databases. The source(s) used will be tailored to the segments and the selected Fellows within these. For example, RePec (Research Publications in Economics) is probably the most well-known economics research database and search engine, and it provides a ranking of individuals and identifies a list of peers who are similarly ranked. It is from this list that a comparison peer for a relevant fellow could be drawn ie if they were working in the field of economics. While PSM could be used to establish the best match, that could lead away from the algorithms databases use to establish similar rankings, so we would instead rely on that, combined with web trawls that aim to identify relevant information in multiple sources.

As illustration we provide an example for Professor Marc Cowling. Marc's primary discipline is economics, so we begin with RePec – the field specialist database. This database

records at monthly intervals all abstract and full paper 'reads', downloads, and citations of Marc's research papers and generates a global, European, and UK overall percentile ranking for him and a more detailed percentile ranking along 33 metrics including 'quality of students'.

A particularly useful tool in RePec is that it provides a list of economic researchers that are similar in quality and general profile to Marc. At this point this whole list of Marc's peers could be used as a relevant match, but for illustrative purposes we selected Michael John Podgursky. The database provides some 33 instantly available research-based metrics available for Michael and Marc all of which are tracked on a monthly basis. All of these metrics would be suitable for statistical analysis and could be used to map the co-evolution of Michael and Marc's research over time if one of them received a particular grant e.g. the FLF and implicitly the other one did not.

However, there is more that can be identified on these two economists, which can be accessed through a Google Scholar search on them both. The front page gives access to life-time research metrics and a record of the last five years. Again this research database can be used to generate additional metrics that are suitable for analysis. Google Scholar is capable of capturing new research in virtually real time even for work that is not formally in the full publication status (for example, 'online first'). Having generated statistical data suitable for comparing Michael and Marc, this source can be used to deepen understanding of their career pathways and the less tangible aspects of their life history. Google Scholar publishes the verifiable homepage for each researcher which can provide the basis for further search and investigation of the matched pair. The information can be compiled into case studies, and trajectories and influence can be compared to consider whether there is evidence that FLF has accelerated the trajectory and outcomes of the matched Fellow.

The search process is relatively quick but does generate a large volume of research and career history information which will take time to process, so we suggest limiting the number of matched case studies to 24. This approach cannot be judged as fully counterfactual due to the potential for unobservable differences between the scholars/pairs involved and there may have been peer researchers not included who would have been a better match. Nonetheless it produces a rich, comparative analysis which may help indicate the degree to which FLF builds the highly influential researchers of the future.

### **7.6.1 Feasibility of processing personal data for the process and theory-based evaluations**

For the process and theory-based evaluations, it is recommended that fellows take part in primary research, namely surveys and qualitative research as well as comparative methods involving secondary data (a matched peer-case study approach). Because of the need to conduct primary research it is important that permissions are in place that enable third parties to make use of personal data to contact individuals for research and evaluation purposes. Any primary research activity would operate on a consent basis, so that while consent (or agreement if public task is used as the rationale for processing) might be in

place that enables contact by an evaluation third party, fellows would still have the option to decline to take part in primary data collection.

The terms and conditions<sup>23</sup> that cover FLF applicants are those of the JeS (Joint Electronic System) which is the on-line electronic system used by seven Research Councils (AHRC, BBSRC, EPSRC, ESRC, MRC, NERC and STFC), to receive funding applications.

As with the counterfactual impact evaluation, the data controller needs to reach decisions about the potential to release personal data to enable contact for primary research. As the process and theory-based evaluations seek only to use data for the 'treatment group' and not for unsuccessful applicants, this is likely to reduce obstacle UKRI foresaw to sharing personal data for unsuccessful applicants to enable contact for primary research.

## 7.7 Summary of recommendations

The process and theory-based evaluation can provide crucial information that the counterfactual impact evaluation is unable to cover. The counterfactual impact will show whether or not the intermediate outcomes were achieved in the fellow's causal pathway, and to a degree the R&I idea pathway, but it will not capture how this was achieved and whether the causal pathways were predictive in the process. It will also not be able to capture the softer, more qualitative outcomes envisaged for fellows. Moreover, as can be seen, the causal pathways of hosts, and EDIP cannot be covered by the counterfactual impact evaluation and it will also not be able to provide information on the partnership as an engine for change (or catalyst) for the intermediate outcomes identified in the theory of change. These are key features of how the FLF is intended to work, which means an impact evaluation alone will be of limited value.

Instead, applying a comprehensive evaluation model that incorporates a range of approaches and methods will mean judgements can be reached on the contributions of the hosts and EDI to the outcomes, as well as explanatory content on the movement through the fellow causal pathway.

Our recommendations for the process and theory-based components of the evaluation of FLF are for:

- **Management information (MI) analysis.** This will document 'inputs' on the four causal pathways related to the agents for change, and capture EDI and porosity indicators at the start of each cohort, from which change can be measured over time. This analysis provides early insight into whether the programme has encouraged and selected a greater range of hosts and PIs involved in MIDRI.
- **Stakeholder and operational interviews.** These interviews will document – from a process perspective – decisions made at each funding round that affect the selection of fellows. Later within the evaluation, interviews focus on strategic assessment against the causal pathways and their predictive ability, for example helping to answer whether gaining funding led to greater certainty and commitment amongst fellows and hosts,

<sup>23</sup> <https://je-s.rcuk.ac.uk/Jes2WebLoginSite/TermsConditions.aspx?mode=accountsetup>



and the reasons for this. Towards the end of the evaluation, strategic stakeholders can address the impact of FLF on the R&I landscape.

- **Fellows – annual, career tracker survey.** This survey series can capture softer measures, attitudinal and behavioural changes to understand whether, for example, the FLF enabled fellows to flourish and thereby to become outstanding researchers who are retained in UK R&I, and then whether and how this leads towards becoming an influential, international R&I leader. It can capture interaction with the EDI pathway and seek to understand more about outcomes for the R&I pathway such as positive competition leading to novel, impactful R&I. It can also explore the ‘engine for change’ as experienced by fellows.
- **Hosts survey.** This survey series, undertaken over three waves, can focus squarely on the host causal pathway, for which there are limited other sources, as well as explore the engine for change further, alongside interactions with the EDI and R&I ideas pathways. It can document the effect of the long-term funding in building confidence to offer better terms and conditions to PIs and whether this trickles down to other R&I staff contracts. Additionally, it can explore the difference made by the FLF in more collaboration and porous arrangements taking root in the R&I sector and whether and how this leads towards the final outcomes envisaged for the programme.
- **Fellow and Host interviews and nested case studies.** This qualitative element will seek to enrich, explain and illustrate the other sources. Interviews can get beneath outcomes to explain how and why they were achieved, and what in the context supported or was an obstacle to outcomes emerging. Nesting interviews so that fellows and hosts are covered means more can be understood about the partnership as an engine for change as the data from these interviews can be triangulated to understand how this led to active and engaged partnerships, that amplified the programme’s aims and helped erode silos to achieve a more porous collaborative system.
- **Fellows and peer matched-cases.** This desk based method would draw on new metrics available due to the transition to digital databases for researchers and research. This would enable quantifiable data to be interrogated for the fellows and matched peers, identified by the key sources in the primary disciplines. Comparative analyses of careers and outcomes (such as number of journals PIs referee for) can help further round out the information on the fellow causal pathway.

## 8 Considerations for economic evaluation

This chapter sets out the considerations for taking forward an economic evaluation or some form of value for money assessment for the FLF. It considers the questions that UKRI wishes to address in this component as well as what this would mean for any evaluation approach.

### Key findings for an evaluation design for FLF:

- The scale, scope and complexity of the FLF programme and its potential benefits as well as the complexity of beneficiaries means that any economic evaluation will require a consideration of multiple potential benefits which may accrue to multiple agents at different levels and at different time. In this sense, the FLF programme may lend itself to separate strands of economic evaluation and CBA.
- In terms of absolute value for money, the fellows' pathways with its robust impact measurement provides one option for exploration, with employment outcomes being possible to measure and value. However, the softer, less measurable outcomes in the fellows' pathways may play an important contributory role to the returns seen to the programme. Moreover, the fellows interact with other causal pathways (hosts and R&I ideas) that will be subject to less robust evaluation, but nevertheless will play an important role in understanding programme costs and benefits.
- Relative value for money will be harder to establish since this implies the need for comparison of cost and benefit data for well-matched programmes or alternative means to establish the same impacts as FLF. While some comparison fellowship schemes exist, they do not cover the full scope of FLF, and they are not subject to the same degree of evaluation. Alternative means to the same impacts as planned for FLF have not yet been established.
- Our recommendation is therefore for a mid-point review within the evaluation timeframe, to take stock of the tangible and intangible information on outcomes that has emerged and to review the potential for value these. Engaging stakeholders in a strategic assessment of the viability of monetising benefits as well as identifying suitable comparators will inform decisions on whether any form of economic assessment is possible.

### 8.1 Approaches to understanding value for money

Economic evaluation focuses on identifying the additional financial value generated through the delivery of an intervention and which can be attributed to that intervention. The additional value is often measured across multiple dimensions and at different points in time. Being able to take forward a robust approach for this, relies on identifying and measuring the additional benefits that derive from an intervention, and valuing those benefits. It then becomes possible to feed these benefits, that are over and above that

which would have been achieved in the absence of the intervention, into a formal Cost-Benefit Analysis (CBA). There are different formats for CBA and so having research questions finalised is important. For example, the interest may be to estimate the return received for every £1 spent on an intervention. Alternatively, the aim may be to compare the ratio of costs to benefits for different programmes to identify the most cost-effective option for delivery.

A cost-benefit analysis can be used to calculate the return on investment from a single intervention, whereas a cost effectiveness analysis is suited to evaluations which seek to test the impact of different interventions. In both cases, the ratio of costs to benefits can be assessed at different levels depending on where costs are likely to be incurred and benefits accrue. For example, some interventions may result in costs or benefits to individuals, employers, society or the exchequer. It can be beneficial to explore the ratio of costs to benefits for different actors as some components may not be apparent at all levels.

The scale, scope and complexity of the FLF programme and its potential benefits as well as the complexity of beneficiaries means that any economic evaluation will require a consideration of multiple potential benefits which may accrue to multiple agents at different levels and at different time. In this sense, the FLF programme may lend itself to separate strands of economic evaluation and CBA.

## 8.2 Understanding the value for money of FLF

The research questions focused on efficiency indicate UKRI's interest in an economic evaluation of FLF.

### Evaluation questions

Based on the overall, estimated impact of the FLF – considering those impacts which can be given market and non-market values – compared to the overall cost of delivering 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?
- To what extent does the FLF represent value for money compared to other possible alternative ways of achieving the same impacts?

Firstly, in absolute terms FLF could be considered to offer 'value for money' if any return from the programme exceeds the amount invested. An analysis of the ratio of cost to the estimated benefits of the FLF programme would provide an insight into whether FLF provides a return on investment.

Secondly, relative value for money is more difficult to assess. To do so it is necessary to identify programmes (or other alternative mechanisms) with a similar impact to FLF and then obtain information on the costs of these interventions. A cost effectiveness analysis can then be used to determine the minimum amount of spending required to achieve a given level of benefit and whether it is more cost-effective to spend money on FLF or on another fellowship scheme (or other mechanism). The review of previous evaluations of similar programmes, summarised in Chapter 4 (and Appendix D), demonstrates that this is

unlikely to be possible, due to the lack of robust evidence on the impact of other fellowship schemes. Also, unless information on the costs of other programmes is in the public domain, it would be difficult to estimate the return on investment that they offer compared with FLF.

In practice, cost effectiveness analysis is generally used when testing competing interventions with similar aims within a single evaluation, due to the difficulty of obtaining comparable information on costs and benefits across different evaluations.

### 8.2.1 Establishing absolute value for money

To achieve this, it would be necessary to subtract the total costs of the programme (ie the public sector 'investment' in the FLF) from a valuation of the net benefit. Net benefits - present values and costs, include the costs of the programme (its funding, staffing and delivery) and further costs relevant to other stakeholders. This is further complicated by time: net benefits from FLF are expected to accrue over a long time period and so a discount rate is used to adjust present values; whereas, costs are incurred over a more limited period and so the costs are not discounted.

In order to calculate a return to investment specifically, it is necessary to know the net impact of FLF across the range of outcomes it seeks to effect and have a credible way of attaching a monetary value to these impacts. However, it may not be feasible to estimate and monetise all possible impacts. For example, taking just one aspect or area of potential impact – which we assess can be established through robust counterfactual methods, building on the impact observed for fellows it could be possible to understand how the receipt of funding has an **impact on fellows' employment** as a proxy for career outcomes, and in turn how this might translate into individual-level earnings. In the economic understanding of efficiency, the programme would then be viewed as efficient if there was a net gain in light of its costs. In this scenario, value for money maximises a positive net benefit to society by minimising the actual costs of the programme, which may include not only programme costs, but also other costs to society. These latter costs are less straightforward to assess since they might include for example, the effect of funding R&I that may result in deadweight, displacement, and substitution, which would have to be considered in the cost calculation.

The monetisation of fiscal benefits would at minimum require an evaluator to monetise increases in taxes/national insurance contributions resulting from additional Gross Value Added ((GVA) (eg measured as work income resulting from the net improved employment created by the programme) in the short and long-term. This would be captured by the net increase in the tax revenue collected by HMRC as an individual receives wage increases over and above those whose starting point was similar due to accelerated career progression.

Further benefits and costs beyond those that are directly measurable include:

1. Non-financial benefits deriving from the receipt of funding eg to quality of life and psychological wellbeing

2. Costs associated with the administration of the programme. This could be proxied by estimating the number of UKRI staff actively engaged in the administration of FLF and a pro rata calculation of the proportion of their time spent on active FLF activity plus some indicator of the seniority level.
3. The EDIP mechanism including valuing the erosion of silos, and planned benefits for increased representation of non-traditional researchers and hosts in the R&I sector.

Once monetised values of the net impacts of FLF are established, a CBA would subtract programme costs from these. Notably, it would be necessary for the CBA to reflect the likely uncertainty in the estimated impacts and monetary valuations by reporting upper and lower bounds.

Employment and earnings progression are important indicators that have immediate relevance to the FLF recipient but these are likely to be strongly correlated to the wider benefits of the FLF programme when the business-case logic model and final outcomes of the TOC are considered. These both focus on impacts on the economy, public good and related to the industrial strategy. Other final and intermediate impacts, such as on an inclusive, international R&I culture and on fellows being inclusive, highly skilled leaders may be challenging to monetise.

Again ‘good’ leadership is likely to be strongly correlated with observable outcomes at the close operating unit level in academia and business, and also at the next hierarchical level up. In this sense good leadership is a relatively intangible input which raises output and efficiency. Moreover, the pathways to these impacts are multiple, with hosts, fellows and the R&I interacting to cause the impacts alongside the effects of EDI and porosity.

Obtaining values for programme costs in terms of public funding may be relatively straightforward for the FLF because monitoring data are likely to be collected consistently to manage the programme and release of funding for R&I. In contrast, obtaining values for the benefits of the programme is far from clear-cut because it involves estimations of all net programme impacts with some, such as being a more inclusive leader, simply unknown.

Due to these uncertainties, at this stage we do not believe it is possible to provide a robust economic evaluation of the programme. However, other components of the evaluation can start to draw together evidence that may inform how this could be addressed as part of summative evaluation.

### **Approaches to establishing cost-effectiveness in other fellowship evaluations**

Other programmes – perhaps as a result of these challenges – have explored different ways to establishing cost-effectiveness and indicate the scale, complexity and relatively early stages of this work. A key example is the evaluation programme of ESRC and its efforts to understand the impact of its work on policy and practice including the economic impact. Their ‘Branching Out’<sup>24</sup> report (2011) summarises a number of different studies

---

<sup>24</sup> <https://esrc.ukri.org/files/research/research-and-impact-evaluation/branching-out-new-directions-in-impact-evaluation-from-the-esrc-s-evaluation-committee/>

which experimented with different approaches to put a value on the impact of ESRC funded work, these are essentially a series of case studies.

The ESRC programme identifies two approaches to economic impact evaluation: tracking forward where the starting point is the research investment, and evaluation attempts to identify impacts that have arisen from the activities; and 'tracking back' from impacts achieved through to investments made to understand returns. This latter approach has been used to measure/quantify the economic impact in the absence of recognised markets for social science outputs, as well as challenges with attribution (e.g. 'disentangling the multiple contributors to policy development') and time-lags. It is less robust than tracking forward but if other options do not exist may provide a way forward. Tracking back stemmed from the conclusion drawn by a contractor 'that robust quantifications of the economic impact of ESRC research must be based on evaluations of the policies to which that research has contributed'. To achieve this, a 'tracking-back' approach was recommended in which the contribution of ESRC-disciplines of research was traced within the development and delivery of a government-level policy evaluation. The ESRC led two pilots to explore the potential of the approach in 2009.

The selected policies were the Education Maintenance Allowance (EMA) and Pathways to Work (PtW) initiatives and two research organisations were contracted to lead the study into assessing the influence of social science on these policies and through a benchmarking process using data from government evaluations of the policies, seek to quantify this influence and then find robust measures of economic value of this influence.

Using EMA as a case example, the contractors found evidence of an interplay between research and the EMA policy, and clear evidence of the contribution to the successful implementation of EMA resulting from a government evaluation. This evaluation resulted in cost efficiencies valued at £80m. The evaluation was undertaken by a number of bodies including an ESRC-funded research organisation, and so the contractor estimated that one eighth of the benefits could reasonably be associated with the ESRC (approximately £10m; which approximated its proportional contribution to the research organisation in relation to the EMA policy). However, the finding was reliant upon an existing evaluation of the policy which established impact and which enabled the impact to be monetarised. Moreover, the findings were heavily caveated given the number of assumptions that were relied upon; as such the method is seen as exploratory rather than fully informative as the potential for deriving robust figures for the scale of funder influence is hampered by the complexities of the routes to influences. It would also be challenging to translate this sort of approach to FLF given its longevity, and range in respect of multi- and inter-disciplinary research and porosity.

## 8.2.2 Establishing relative value for money

To derive a value for money compared to other ways to achieve the same impacts will also be challenging and would require primary research to understand views, once FLF programme impacts are established, on what could or has achieved the same impacts as FLF. A query surrounds whether alternative fellowships should be the source of contrast; in which case a comparative analysis would rely on the same information being captured on costs as well as benefits as is for FLF. The work to date in the feasibility study has



established that there are some fellowship schemes that could provide a useful comparison on outcomes (namely the ERC Starting Grants and the Wellcome Trust/Royal Society Sir Henry Dale Fellowships) but further work would be required to understand if cost and benefit data is being collected that would enable comparison on the same dimensions as FLF. The comparability of inputs and outputs would need to be established alongside cooperation to share cost data to understand differential returns between programmes.

The feasibility study has not, however, focused on alternative, non-fellowship means to achieve the impacts expected for FLF, and these impacts are complex in many ways. Hence, understanding if there was a different way to fund R&I to have an impact related to the industrial strategy may be possible, but to understand objectively alternative means to create inclusive, highly skilled international R&I talent would be more challenging.

## 8.3 Summary recommendations

It is clear that the scale, scope and complexity of the FLF programme including of the potential benefits and beneficiaries requires detailed consideration to formulate how an economic assessment could be conducted. The programme has some unique features which require understanding, and where it shares commonalities with other programmes, these have not necessarily been subject to robust economic, cost-benefit or value for money assessments. The scoping work shows how multiple potential benefits may accrue to multiple agents at different levels at different timescales. In this sense, the FLF programme may lend itself to separate strands of economic evaluation and CBA although, it is also apparent that the FLF programme does not easily lend itself to benchmarking against other evaluations in its totality. Rather, a more disaggregated approach is required that identifies separate strands of evaluation and CBA, for example, at the individual fellow level, their immediate working group, departmental or operating division level, the university or firm, and also to wider society – this again reflects the complexity of the programme.

We have also established that any economic assessment process must operate along multiple dimensions and that the timing of any measurement of additional benefits is important with some benefits being derived in a short time-frame and some taking many years to be fully maximised. It follows that the precise timing of such an assessment (from the time at which the fellowship was awarded) should inform the most relevant outcome and benefits measurement and indicators. On the latter when outcome and benefit measures can be monetised, then an appropriate monetary deflator should be used to adjust for the time value of money.

As is the case in all evaluations there are naturally 'hard' and 'soft' indicators both of which have relevance. Hard indicators lend themselves to precise measurement and are more tangible, for example growth in employment. They also lend themselves to formal econometric evaluation techniques. Soft indicators are less tangible and often seek to measure behavioural changes or in 'quality' of leadership. But they are often a key element in the causal evolution of harder benefits and should not be ignored simply because they are difficult to identify, measure and quantify. These aspects of any evaluation process lend themselves to qualitative investigation. However, even in a formal CBA these softer



indicators can be included as an important addition even when a monetary value cannot be easily assigned.

This range of uncertainty and limited information indicates that any evaluation must first seek to understand and document the action of FLF to provide information for how an economic assessment could be drawn and the best approaches to address this. Our recommendation would be for a mid-point review within the evaluation timeframe, to take stock of the tangible and intangible information on outcomes that has emerged. Engaging stakeholders in a strategic assessment of the viability of monetising benefits as well as identifying suitable comparators will inform decisions on whether any form of economic assessment is possible.

## 9 Recommendations for a full evaluation

This chapter covers the recommendations for an evaluation of the FLF, following the detailed work to scope the potential and consider the feasibility of different approaches. It addresses the aims and objectives set for the scoping and feasibility work, and shows how the evaluation approach recommended will meet the aims for the study of FLF.

### Key issues for the FLF evaluation

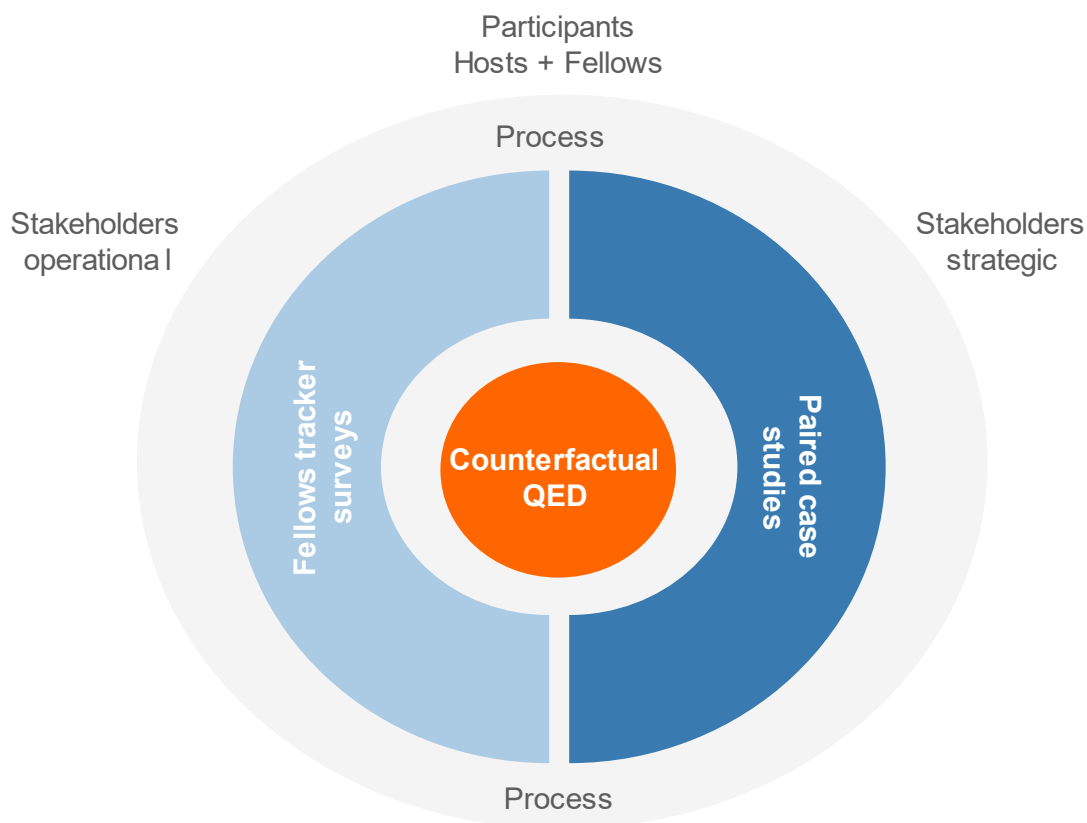
- The research questions set for scoping and feasibility study indicated that it was necessary to establish if the FLF could be subject to robust evaluation, if so, what approaches would best address this, what could measure impact, at what level, and over what time-scale.
- The research indicates that a comprehensive, multi-component evaluation can be used that will deliver evidence on impact with varying levels of robustness. The evaluation should apply: net impact, theory-based impact; process and strategic assessment elements.
- The most robust, counterfactual net impact measures can be applied at the individual fellow level. These will reveal whether the programme helped to create a more inclusive, international R&I culture through supporting new inclusive, influential PIs to emerge. This will also provide insights related to the R&I and host causal pathways in the theory of change.
- Theory-based impact evaluation will track the causal pathways of fellows, hosts, the R&I and EDIP, confirming and elaborating the interaction of these in achieving the planned final impacts of ground-breaking R&I with positive outcomes for society, the economy and the UK's R&I sector.
- Process evaluation at the operational level will chronicle the development and delivery of the programme, ensuring that decisions that affect inputs and activities are captured so the effects can be accounted for in the impact studies. At a strategic level, senior stakeholders can assist the assessment of the impact of FLF on the UK research landscape.
- A judgement has been reached that it is too soon to know if an economic evaluation can be embedded and what form this can take. The recommendation is a mid-term review is undertaken, engaging with strategic stakeholders, that can identify whether the emergent outcomes can be credibly valued.

The overarching aim of this scoping and feasibility study was to identify the appropriate methods to support answering the FLF evaluation questions by proposing a suitable, workable approach to monitoring and evaluating the FLF programme. The objectives were to summarise existing knowledge and information regarding effective monitoring and evaluation of fellowship instruments; explore the challenges of the FLF evaluation and discuss appropriate methods to address them; and recommend a suitable, workable approach to monitoring and evaluating the FLF.

This extensive review of the potential to evaluate the FLF programme has identified that the application of a comprehensive evaluation model encompassing counterfactual and theory-based impact, process and strategic assessment of the programme, will be the optimal approach to understanding the programme's effects at the individual and aggregate level. The theory of change identified a complex programme with multiple causal pathways which should act in an additive way to achieve the final impacts on society and the economy. It showed the fellows and hosts are both intermediaries – alongside porosity/MIDRI and EDI (summarised as EDIP) – to the final impacts.

Due to the complexity of the programme, a comprehensive evaluation is necessary because there is a need to explain not only if the programme had the planned impact, but how it had this impact. Furthermore, the robustness of the measures that are available to understand anticipated impacts vary. As a result, the programme will benefit from the combination of a full counterfactual impact evaluation; theory-based impact evaluation; and process evaluation.

**Figure 9-1: Overview of evaluation model**



Source: IES, 2021

The following methods are proposed:

- Full counterfactual impact evaluation using a quasi-experimental design (QED). This component of evaluation, which traces the fellow's causal pathway, will focus on

programme applicants and use unsuccessful applicants as a counterfactual group for grant recipients. This will allow the impact of receipt of FLF funding over not receiving this funding to be captured. While it focuses on the individual fellow, because of the inter-relationship with hosts and R&I ideas, there will be insights for these causal pathways too (although net impact will be established at the individual level). Two approaches are proposed: propensity score matching based on the whole applicant group; and a Regression Discontinuity Design based on those unsuccessful applicants who were shortlisted to interview. Outcomes will be measured relative to information collected at baseline (application) and then in secondary, administrative data. Data is available that can measure outcomes at the individual fellow and host level. It must be noted however, that the likelihood is that this approach will be most viable for those fellows in academia – as the FLF receives most applications and makes more awards in the segments and only small numbers in the business segments. Nonetheless, the prospects for drawing comparison groups in the unsuccessful applicant data should be fully explored once funding decisions have been made for cohorts 5 and 6.

- Theory-based impact evaluation will enable comparison between FLF and other programmes (the impact of FLF relative to other schemes) as well as a closer assessment of the causal pathways to impact for hosts and the R&I (including the ‘engine for change’ contained within the partnership at the heart of FLF). It will enable investigation of the softer outcomes for leadership and in relation to EDIP. This will complement the counterfactual impact evaluation by providing insights into how outcomes are emerging for the fellow pathway, and the interplay with the other pathways to impact. As such, this component will focus on the fellows, who will be involved primary data collection through surveys and interviews; and hosts who will be invited to participate in surveys. Qualitative nested case studies (focused on the fellow, host and R&I) will provide exemplars to how the partnership is an engine for change and goes on to create the pathways to impact. Surveys will use common measures with other programmes so there is potential to benchmark outcomes. An innovative, peer matched-case approach will track a sample of fellows against matched peer researchers drawn from secondary data. Interviews will stakeholders will identify the impact of the programme on the R&I landscape in the UK.
- Process evaluation will track and document the development and delivery of the programme. Inputs and activities will be captured including decisions made at each funding point, and how this affected the fellows and projects that were selected for funding. It will then trace the support offered to fellows by the UKRI through the dynamic development programme and fellows’ views of this. The approach will be predominantly qualitative involving operational stakeholders, although the fellows career tracking survey will gather information on programme experiences and views of the support available. This component will analyse and report the baseline MI.

## 9.1 What this means for the scoping and feasibility questions

The key questions for the scoping and feasibility study were:

## 1. Is it possible to provide a robust evaluation of the FLF, and if so when would be the most appropriate time to instigate this?

The review has established that the degree of robustness will vary within the FLF evaluation. Nonetheless, it is possible to take forward a suite of approaches that will identify the range of impacts stemming from this complex programme.

- A full, counterfactual net impact evaluation can be used to address questions of impact at the individual, fellow level. This will also provide insights into inter-related causal pathways of (i) the R&I idea which is the key generator for the final impacts on society, the economy and industrial strategy and (ii) the host since the fellow's activities and progress will be enablers – alongside the funding - for the host outcomes to emerge.
- Theory-based impact evaluation – a robust approach that is recognised in the Magenta Book – can be used to detect progress along the causal pathways in terms of outcomes, alongside attitudinal and behaviour changes, to understand the effects of FLF more roundly on Fellows (particularly the softer outcomes expected), and importantly for hosts. Theory-based evaluation can scrutinise the effect of EDIP outputs and outcomes and whether these indicate final impacts on a more inclusive, international R&I culture will emerge, and trace the pathway of the R&I idea(s).
- The FLF is a large, complex programme seeking to effect change on a range of measures at the societal level. These will take significant time to emerge (in all likelihood). The recommendation from this review is to embed an evaluation that can continue beyond the funded period, ideally for a minimum of six years. This will allow more granular insights into impacts for different disciplines and MIDRI mixes, as well as in different host settings (related to the recommended segmentation).
- With these points in mind, a review of the intake following the award of round 6 will mean the viability of taking forward the most robust impact measurement for all fellows or a subset related to the four segments related to host and discipline can be confirmed.

## 2. What is the most appropriate broad, sensible and feasible approach to evaluating the FLF scheme

- i. How do measurements relate to a theory of change?
- ii. How often should they be captured?
- iii. At what level should measurements be made?

To be able to capture the full operation of the programme, a multi-component evaluation is required. This will enable UKRI to understand (i) whether or not the programme achieved its planned impacts and (ii) for whom and in what circumstances did it achieve this and (iii) what factors in the context or operation of FLF supported or inhibited the emergence of impacts. A range of measures have been identified in administrative sources that alongside primary data collection can plausibly cover the effects anticipated for FLF to the intermediate impact level at minimum.

- The measures have been selected to capture outcomes identified in the theory of change – mainly at the intermediate impact level for the fellow, host, and EDIP

pathways that will provide insights to the R&I pathway - while keeping in mind the need to minimise burden on the sector and to measure in line with DORA principles. Use of administrative data sources not only minimises burdens but increases the chances of collecting outcomes for a large proportion of fellows, since this overcomes the challenges of response rate and attrition that are seen in primary, survey-based data collection.

- Primary data collection remains necessary to fully round out the picture for fellows, and to get to host outcomes, where administrative sources are more limited.
- Attitudinal and behavioural changes are particularly important to the fellow pathway and link strongly to the EDIP pathway. Engaging fellows in annual survey research through a career tracker will capture these outcomes and help establish how they lead (or do not lead) to the planned intermediate outcomes.
- Host surveys are crucial, however are less necessary on an annual basis, and would create undue burden. Capturing information at baseline, at the end of the initial funded period, then following the contract extension period is likely to be optimal.
- Given the complexity of the programme, measures must encompass the four key segments related to host and discipline, and must be sensitive enough to pick up on MIDRI discipline mixes. The key unit of assessment for the counterfactual impact evaluation is the individual fellow. The theory-based evaluation can go beyond this to encompass hosts, EDIP and R&I and effects in aggregate.

### 3. Given the balance of academic/business-based fellows, how should this be reflected in any potential evaluation?

The balance of academic and business-based fellows has been a key consideration in the scoping study. The theory of change set out the ambitions that the FLF encourages a greater range of host organisations into R&I (at the theory of change input level) and that this will create greater diversity, that hosts will be retained in the sector increasing the extent of positive competition and collaboration, and that in turn will lead onto system that supports ground-breaking R&I.

The reality of FLF is that awards have been made predominantly to hosts in academia and the businesses are a small group of hosts. This will limit the ability to find granular counterfactual impact for fellows (there may be too few good matches in the unsuccessful applicants to enable comparison), and it presents risks in terms of business hosts not responding to primary research in the theory-based evaluation. Using the matched case, theory-based method, drawing on publication databases, will help to surface the impacts for these smaller segments if robust methods do not prove viable.

### 4. How, and to what extent, can the different impacts of the FLF be measured (both objectives from the business plan, and possibly also softer impacts of leadership and behavioural change)?

The scoping and feasibility study has established the means through which the hard outcomes and the softer ones can be measured. It is essential to attempt to measure both as key aspects of the theory of change could not be tested without this. For example, the EDIP causal pathway envisages that fellows not only will be diverse and non-traditional, but that they will value this and this will affect their own (and their host's) behaviour in respect of building research teams and collaborating. It is through breaking down silos that ambitious R&I, will lead onto leading edge R&I and then onto societal and economic impacts.

The approach proposed here will capture hard metrics for the fellows that are consistent with R&I but stretch beyond the narrow measures that are often used. These are drawn from administrative data sources. Some of this information can be used to understand PI behaviour around collaboration for example, so while it might be considered 'hard' in focus it can also contribute to understanding softer behaviours. Primary research with fellows and hosts will add dimensionality to this, and allow the interaction between the inputs, and outcomes to be better understood to firm up the mechanisms through which the FLF is leading to impact.

There are limitations and constraints for the evaluation that extend beyond the methodological and instead relate to practical decisions made about evaluation. For example, it is likely that only long-term evaluation will establish the effect of FLF at the major/final impact level envisaged in the theory of change; and the extent to which granular detail on the causal pathways and interaction between them will depend on the frequency of evaluation rounds, and extent of engagement in each.

## 9.2 The design of the FLF evaluation

The evaluation framework (Table 9-1) below shows how the recommended components and range of methods for evaluation interact to fully address the research questions posed by UKRI, with the exception of the economic evaluation where further scoping and benchmarking is required to establish a credible approach to establishing cost-effectiveness. A mid-term strategic assessment involving senior stakeholders will be crucial to this element.

**Table 9-1: Evaluation framework**

Research question	Method	Issues
<b>Appropriateness questions – Process evaluation</b>		
<b>To what extent (and how) is the FLF programme working and being delivered as intended?</b>		
(To what extent) are targets for FLF inputs and outputs being met?	<ul style="list-style-type: none"> <li>• MI analysis</li> <li>• Stakeholder interviews</li> </ul>	Fund value, and level of funds distributed to fellows in each of the established segments No. of fellows in each round and overall
(To what extent) has FLF stimulated multi- and interdisciplinary research and	<ul style="list-style-type: none"> <li>• MI analysis</li> <li>• Stakeholder interviews</li> <li>• Case study exemplars</li> </ul>	Characteristics of fellows, hosts and projects in each round and overall



Research question	Method	Issues
innovation (MIDRI) fellowships?	<ul style="list-style-type: none"> <li>Peer matched case method</li> </ul>	
(To what extent) are FLF fellowships additional to other schemes supported by UKRI and other UK / international funders?	<ul style="list-style-type: none"> <li>Stakeholder interviews</li> <li>Comparative analysis against existing evidence</li> <li>Career tracker survey</li> </ul>	<p>Views and evidence on uniqueness of FLF contribution</p> <p>Common measures in career tracker surveys to benchmark to other programmes/funding schemes</p>
(To what extent) has FLF delivered effective post award management to support the professional development of the fellows?	<ul style="list-style-type: none"> <li>Programme documentation</li> <li>Career tracker survey</li> <li>Qualitative interviews with fellows</li> <li>Stakeholder interviews (operational)</li> <li>Peer matched case method</li> </ul>	<p>Content of the programme over time</p> <p>Views on the composition and effectiveness of the dynamic support programme</p>
What, in practice, is felt to be working more / less well regarding the delivery of the fund, and why?	<ul style="list-style-type: none"> <li>Stakeholder interviews (operational and strategic)</li> <li>Qualitative interviews with fellows</li> <li>Qualitative interviews with hosts</li> </ul>	Views on delivery – what is working well, what is less effective
What are the unexpected barriers or facilitators to the FLF processes and the delivery of the anticipated outcomes, if any?	<ul style="list-style-type: none"> <li>Stakeholder interviews (operational and strategic)</li> <li>Qualitative interviews with fellows</li> <li>Qualitative interviews with hosts</li> </ul>	Views on delivery – what in the context is supporting/preventing the achievement of outcomes
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?	<ul style="list-style-type: none"> <li>MI analysis</li> <li>Stakeholder interviews</li> <li>Career tracker survey outcomes</li> </ul>	<p>Descriptive analysis of who (hosts and fellows) reached by FLF</p> <p>Views and interpretation on programme targeting and implications</p>
What lessons are there for future rounds / similar schemes?	<ul style="list-style-type: none"> <li>Stakeholder interviews</li> <li>Meta/comparative analysis</li> </ul>	<p>Benchmarking to other schemes</p> <p>Views on lessons gleaned by key stakeholders</p>
<b>Effectiveness questions – Impact evaluation</b>		
The evidence base for the impact evaluation questions will develop progressively over the near, mid and long term.		
<b>(How) has FLF changed the research and innovation landscape?</b>		
(To what extent) has FLF increased high quality and impactful research and innovation?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Strategic stakeholder interviews</li> </ul>	<p>Evidence on FLF impact over not receiving FLF</p> <p>Views on the additionality of the FLF</p>
(To what extent) and how does the overall FLF fund	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Strategic stakeholder interviews</li> </ul>	Evidence on FLF impact over not receiving FLF

Research question	Method	Issues
support wider government objectives?	<ul style="list-style-type: none"> <li>Theory-based impact nested case studies</li> </ul>	<p>Views on the additionality of the FLF over other schemes related to government objectives</p> <p>Exemplars of how FLF made this difference</p>
(To what extent) has FLF increased MIDRI and cross-sector working, for the Fellow, Fellow's team and / or collaborators?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellows and host surveys</li> <li>Theory-based impact matched peer case studies</li> </ul>	<p>Evidence on FLF impact over not receiving FLF</p> <p>Evidence of attitudinal and behavioural changes to porosity, MIDRI, and collaboration</p> <p>Comparative evidence on propensity to be engaged in collaborative, MIDRI research</p>
<b>To what extent, and how, has FLF delivered highly skilled research and innovation leaders of the future?</b>		
(To what extent) has FLF developed R&I leaders over and beyond other fellowship (or similar) schemes?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellows surveys using common metrics with other schemes</li> <li>Theory-based impact matched peer case studies</li> </ul>	<p>Evidence on FLF impact over not receiving FLF</p> <p>Evidence of leadership behaviours aligned to programme aims and application benchmarked against other schemes</p> <p>Comparative evidence based on similar peers</p>
<b>To what extent, and how, does FLF make the UK attractive place for future R&amp;I leaders?</b>		
(To what extent) has FLF developed attracted and retained talent (fellows and associated teams) to the UK?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic stakeholder interviews</li> </ul>	<p>Evidence of impact of FLF on retention over not receiving FLF</p> <p>Views on likelihood of continued engagement in UK-based R&amp;I</p> <p>Views of specific and additional effects of FLF</p>
To what extent, and how, has FLF influenced the reputation of the UK as a place to pursue a career in research or innovation?	<ul style="list-style-type: none"> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic stakeholder interviews</li> </ul>	<p>Views on role of FLF in influencing decisions about where to pursue an R&amp;I career</p> <p>Views on the specific and additional effects of FLF, in context of wider change over time</p>
<b>(How) has FLF led to a change in behaviour for early career researchers, innovators and hosts?</b>		
(To what extent) has FLF increased engagement between industry and academia on research and innovation activities?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellow and host surveys</li> <li>Theory-based impact matched peer case studies</li> </ul>	<p>Evidence of impact of FLF on retention over not receiving FLF</p> <p>Views and behaviours related to collaboration and porosity</p>

Research question	Method	Issues
		Comparative evidence on propensity to engage more frequently in industry/academic collaborative endeavours
(To what extent) has FLF developed a more equal, diverse and inclusive research and innovation workforce?	<ul style="list-style-type: none"> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic and operational stakeholder interviews</li> </ul>	Views and behaviours centred on R&I workforce EDI Views on the extent to which FLF contributed to a more inclusive workforce; factors in the context (also) affecting this outcome
To what extent, and how, has FLF influenced the appetite to risk to novel R&I of early career researchers and innovators, panel members and hosts?	<ul style="list-style-type: none"> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic and operational stakeholder interviews</li> <li>Theory-based impact nested case studies</li> </ul>	Views and behaviours centred on risk related to novel R&I and ECR&Is Views on the additionality of FLF on this outcome vs other factors in the context Exemplars of how FLF supported this
(To what extent) have host organisations promoted and supported the FLF scheme and delivered against expectations / commitments for research or innovation support?	<ul style="list-style-type: none"> <li>Theory based measures in host and fellow surveys</li> <li>Theory-based impact nested case studies</li> <li>Strategic and operational stakeholder interviews</li> </ul>	Views on the degree to which long-term sustained funding supported this
(To what extent) has FLF increased careers in research and innovation within new and novel areas?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellow and host surveys</li> <li>Theory-based impact matched peer case studies</li> </ul>	Evidence of impact of FLF on careers over not receiving FLF Views on the extent of contribution on this theme Comparative examples of difference made
(To what extent) has FLF contributed to new investment into research and innovation from outside government?	<ul style="list-style-type: none"> <li>QED counterfactual impact</li> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic stakeholder interviews</li> </ul>	Evidence of impact of FLF on careers over not receiving FLF Evidence on new inwards investments to hosts and fellows Views on the additionality of FLF over other schemes and contextual factors affecting this outcome
(To what extent, if any) has FLF influenced, or set precedents for, improvements in Host Organisations' support for early career researchers or	<ul style="list-style-type: none"> <li>Theory-based measures in fellow and host surveys</li> <li>Strategic stakeholder interviews</li> </ul>	Views and behaviours on improvements for support offered to new R&I talent Views on the additionality of FLF over other schemes and

Research question	Method	Issues
innovators, EDI or related UKRI policy goals?		contextual factors affecting this outcome
<b>To what extent (and how) has the FLF delivered wider knowledge, economic and societal impacts?</b>		
What has been the wider, overall impact of the FLF on UK research and innovation expertise and on other parts of UKRI practice?	<ul style="list-style-type: none"> <li>• QED counterfactual impact</li> <li>• Theory-based measures in fellow and host surveys</li> <li>• Strategic stakeholder interviews</li> </ul>	<p>Evidence of impact of FLF on careers over not receiving FLF including spillovers</p> <p>Views and examples of spillover effects from FLF</p> <p>Views on effects on UKRI practices and UK R&amp;I system</p>
What has been the wider, overall economic impact of the FLF, including the economic value of non-market impacts?	<ul style="list-style-type: none"> <li>• Strategic stakeholder interviews</li> <li>• Theory-based measures in fellow and host surveys</li> <li>• QED counterfactual impact</li> </ul>	<p>Views on the additionality of FLF over other schemes and contextual factors affecting this outcome</p> <p>Evidence and views on economic effect on hosts and R&amp;I talent</p> <p>Evidence of impact of FLF over not receiving FLF</p>
What has been the wider, overall societal impact of the FLF, including the impact on the number of high-quality jobs, wages, and wellbeing, and societal benefits from new products and services?	<ul style="list-style-type: none"> <li>• Strategic stakeholder interviews</li> <li>• Theory-based measures in fellow and host surveys</li> <li>• Theory-based impact matched peer case studies</li> </ul>	<p>Views on the effect and additionality of FLF over other schemes and contextual factors affecting this outcome</p> <p>Views on difference made by FLF in relation to these measures</p> <p>Evidence on employment tenure and nature based on similar comparator cases</p>
<b>Efficiency questions – Economic evaluation</b>		
It is acknowledged that it will be difficult to derive a robust assessment of value for money based on an estimate of the value of impacts. Nonetheless, we are keen to understand...		
<b>Based on the overall, estimated impact of the FLF – considering those impacts which can be given market and non-market values – compared to the overall cost of delivering the FLF, to what extent does the FLF represent value for money?</b>		
To what extent does the FLF represent value for money in absolute terms?	<ul style="list-style-type: none"> <li>• Mid-term review</li> <li>• Strategic stakeholder interviews</li> <li>• Economic assessment would be based on valuing outcomes derived from QED counterfactual impact and cost information</li> </ul>	<p>Views on the additionality of FLF; views on the value of benefits; views on costs</p> <p>Gathering full cost data on programme delivery and development</p> <p>Inferring value of impacts based on similar schemes/mechanisms</p>
To what extent does the FLF represent value for money compared to other possible	<ul style="list-style-type: none"> <li>• Mid-term review</li> <li>• Strategic stakeholder interviews</li> <li>• Economic assessment would be based on valuing outcomes</li> </ul>	Views on the differences/similarities and additionality of FLF over other schemes

---

Research question	Method	Issues
alternative ways of achieving the same impacts?	derived from QED counterfactual impact and cost information, and comparing to evidence on economic evaluation other schemes/ mechanisms	Benchmarking value of impacts based on similar schemes/ mechanisms; benchmarking costs to other schemes/ mechanisms

## References

---

- AHRC (2009). Leading the world: the economic impact of UK arts and humanities research, Arts and Humanities Research Council.
- Barr, J., Bryan, B., Kolarz, P., Potau, X., Puncton, M., Simmonds, P. and Vogel, I. (2018). Global Challenges Research Fund Evaluation: Foundation Stage, Final Report, e-Pact, Oxford.
- Befani, B. (2016) Choosing appropriate evaluation methods: a tool for assessment and selection, Bond.
- BiGGAR Economics (2019). Royal Society of Edinburgh Enterprise Fellowships - Evaluation and Economic Impact Assessment, Scotland: BiGGAR Economics.
- Bryer, N., Bebb, H. & Evans, N., (2018). Sêr Cymru II, Mid-Term Evaluation, Cardiff: Ymchwil OB3 Research and Regeneris for the Welsh Government.
- BEIS (2019). International comparative performance of the UK research base, 2019: Accompanying note, Department for Business, Energy & Industrial Strategy (BEIS).
- BEIS/DfE (2020). Reducing bureaucratic burden in research, innovation and higher education, policy paper.
- BEIS/Innovate UK (2017). Catapult programme: a framework for evaluating impact.
- BIS (2012). Leadership & Management in the UK – the key to sustainable growth: a summary of evidence for the value of investing in leadership and management development, Department for Business, Innovation & Skills, 12/923
- Cabinet Office (2021). Global Britain in a competitive age: the integrated review of security, defence, development and foreign policy.
- CRAC and IES (2020). Royal Academy of Engineering Research Programmes: Evaluation and Career Tracking, Cambridge: Report by the Careers Research & Advisory Centre (CRAC) supported by the Institute for Employment Studies (IES) for the Royal Academy of Engineering.
- Cefis, E., & Orsenigo, L. (2001). 'The persistence of innovative activities: A cross-countries and cross-sectors comparative analysis'. *Research Policy*, 30(7), 1139– 1158.
- CBR/PACEC (2009). Evaluation of the effectiveness and role of HEFCE/OSI Third Stream Funding. A report to HEFCE by PACEC and the Centre for Business Research, University of Cambridge.

Coad, Cowling, Nightingale, Pellegrino, Savona, and Siepel, (2014). UK Innovation Survey – Innovative Firms and Growth, Report to DBEIS.

Coates Ulrichsen, T. (2015) Assessing the economic impacts of the Higher Education Innovation Fund: a mixed-method quantitative assessment, a report for HEFCE, Higher Education Funding Council England (HEFCE).

Cohen, W. (2010). Fifty years of empirical studies of innovative activity and performance. In Hall B. and Rosenberg N. Handbook of the Economics of Innovation. Vol. 1 (Vol. 1, pp. 129–213). North-Holland.

Couerduroy, R., Murray, G., Cowling, M. and Licht, G. (2012) Adolescent Firm Survival and Internationalisation. International Small Business Journal, 30 (5). August. 472-492.

Davenport. A, and Zaranko, D. (2020). Levelling up: where and how?, IFS Green Budget 2020: Chapter 7, Institute for Fiscal Studies.

D’Este, P., Iammarino, S., Savona, M., & Tunzelmann, N. von. (2012). What hampers innovation? Revealed barriers versus deterring barriers. Research policy, 41(2), 482–488.

ESRC (2009). Taking stock: a summary of ESRC’s work to evaluate the impact of research on policy and practice, Economic & Social Research Council.

ESRC (2011). Branching out: new directions in impact evaluation from the ESRC’s Evaluation Committee, Economic & Social Research Council.

ESRC (2013). Cultivating connections: innovation and consolidation in the ESRC’s impact evaluation programme, Economic & Social Research Council.

Elsevier (2016). International comparative performance of the UK research base 2016, Department for Business, Energy & Industrial Strategy (BEIS).

Elsevier (2013). International comparative performance of the UK research base - 2013, Department for Business, Innovation and Skills (BIS).

Elsevier (2011). International comparative performance of the UK research base - 2011, Department for Business, Innovation and Skills (BIS).

European Commission, (2017). FP7 ex post and H2020 interim evaluation of Marie Skłodowska-Curie actions (MSCA), Luxembourg: Publications Office of the European Union.

European Commission (2017). Commission staff working document: interim evaluation of Horizon 2020. Luxembourg: Publications Office of the European Union.

Freshney Consulting and Aleron,(2019). Review of the Springboard Scheme for Biomedical Scientists, s.l.: Freshney Consulting and Aleron for the Academy of medical Sciences.

Gleed, A. and Bennett, R. (2018). NERC Science or the Environment, Evaluation of support for ECRs, Cheshire: djs research.



Grant, J., Brutscher, P-B., Guthrie, S., Butler, L. and Wooding, S. (2010). Capturing Research Impacts: A review of international practice, Documented Briefing, RAND Corporation.

Guerrero-Bote, VP., Chinchilla-Rodríguez, Z., Mendoza, A. and de Moya-Anegón, F. (2021). Comparative Analysis of the Bibliographic Data Sources Dimensions and Scopus: An Approach at the Country and Institutional Levels. *Front. Res. Metr. Anal.* 5:593494. doi: 10.3389/frma.2020.593494.

Handforth, R. and Mellors-Bourne, R. (2020) Royal Academy of Engineering Research Programmes: Evaluation and Career Tracking, Royal Academy of Engineering.

HM Government (2020). UK research and development roadmap.

HM Treasury (2020). Spending Review 2020, CP330.

HM Treasury (2020a). Magenta Book: Central government guidance on evaluation.

Hodges, D. (2018). Setting the Scene: How the UK Supports Innovation, Innovate UK.

Hopkins, M. M., & Siepel, J. (2013). 'Just how difficult can it be counting up R&D funding for emerging technologies (and is tech mining with proxy measures going to be any better)?' *Technology Analysis & Strategic Management*, 25(6), 655–685. doi:10.1080/09537325.2013.801950.

Hornbostel, S. et al., (2009). Funding of young scientist and scientific excellence. *Scientometrics*, Volume 79, p. 171–190.

Hutton, W. and Nightingale P (2011). The Discouraged Economy. Work Foundation London.

Industrial Strategy Council (2019). Measuring the success of the industrial strategy: schedule of success metrics.

Innovate UK (2018). Evaluation Framework: How we assess our impact on business and the economy, Technology Strategy Board, Innovate UK, Swindon

Ipsos MORI, George Barrett & Tomas Ulrichsen, (2018). ICURe Evaluation Report, London: Ipsos MORI, Social Research Institute.

Jenkins, S. and Bryant, C., (2012). Clinician Scientist Fellows Scheme Evaluation 2012, London: Jenesys Associates for the Academy of Medical Sciences.

Jones, M.M., Manville, C. and Chataway, J. (2017) Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future, *J Technol Transf*, Springer.

KADA Research, (2020). Interim Evaluation of the DigitalHealth.London Accelerator, Sheffield: KADA Research.

King's College London and Digital Science (2015). The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies, Research Report 2015/01, HEFCE.

Kotvojs, F. and Shrimpton, B. (2007). 'Contribution analysis: A new approach to evaluation in international development', *Evaluation Journal of Australasia*, 7(1), p27-35.

Lööf, H. Johansson, D, Andersson, M. and C Karlsson (2012). R&D Strategy and Firm Performance, What is the Long-run Impact of Persistent R&D?, Chapter 8 in Andersson, M., Johansson, B., Karlsson, C., & Lööf, H. (2012). *Innovation and Growth: From R&D Strategies of Innovating Firms to Economy-wide Technological Change*. Oxford University Press, Oxford.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E. et al. (2021) 'Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations'. *Scientometrics* 126, 871–906. <https://doi.org/10.1007/s11192-020-03690-4>.

Martinez, A., Epstein, C.S. and Parsad, A. (2016). 'Developing internationally engaged scientists and engineers: The effectiveness of an international postdoctoral fellowship program'. *Research Evaluation*, 25 (2), 184–195.

Meagher, L., (2004). Evaluation of the Postdoctoral Fellowship Scheme, s.l.: Technology Development Group, Research Evaluation Committee, ESRC.

Mellors-Bourne, R. and Metcalfe, J. (2015). Principal Investigators and Research Leaders Survey (PIRLS): 2015 UK aggregated results, Vitae/Careers Research & Advisory Centre, Cambridge

Mellors-Bourne, R., Metcalfe, J. & Pollard, E., (2018). Royal Society Research Fellowships: Career Pathway Tracker, Cambridge: Careers Research & Advisory Centre (CRAC) with support from the Institute for Employment Studies (IES).

Meyer, N. & Bühner, S., (2014). Impact Evaluation of the Erwin Schrödinger Fellowships with Return Phase, Karlsruhe: Fraunhofer.

NCUB (2016). Latest collaboration progress monitor published, National Centre for Universities and Business.

NAO (2021). UK Research and Innovation's Management of the Industrial Strategy Challenge Fund, Report by the Comptroller and Auditor General, Session 2019-2021 HC 1130.

Nedeva, M., et al (2012). Understanding and assessing the impact and outcomes of the ERC and its funding schemes: final synthesis report, European Research Council

NESTA (2009). The Innovation Index: measuring the UK's investment in innovation and its effects, Index Report, NESTA.

Nurse, P. (2014). Ensuring a successful UK research endeavour: A review of the UK Research Councils, BIS/15/625, Department for Business, Innovation and Skills, London.

O'Flynn, P. and Barnett, C. (2017). Evaluation and impact investing: a review of methodologies to assess social impact, Evidence report 222, New Frontiers in Evaluation.

- Park, T., Zhou, M., Salstrom, R., & Yi, J. (2018). 'Patenting and Licensing Decisions: A Tested Method for Practical Decisions'. *Journal of Supply Chain and Operations Management*, 16(2), 176.
- Parks, S., Rodriguez-Rincon, D., Parkinson, S. and Manville, C. (2019). The changing research landscape and reflections on national research assessment in the future, RAND Europe.
- Pattyn, V., Molenveld, A. and Befani, B. (2019). 'Qualitative comparative analysis as an evaluation tool: lessons from an application in development cooperation', *American Journal of Evaluation*, 40(1), p55-74.
- Rogers, P.J. (2000). 'Causal Models in Program Theory Evaluation', in P. J. Rogers , A. J. Petrosino , T. Hacsí and T. A. Huebner (eds) *Program Theory Evaluation: Challenges and Opportunities*, pp. 47—55, New Directions in Evaluation series. San Francisco, CA: Jossey-Bass .A
- Siepel, J., Cowling, M., & Coad, A. (2012). Long run growth and survival of UK high tech firms. London: NESTA.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012). Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development, working paper 38, Department for International Development (DfID).
- Stewart, P. et al., (2012). 'What happens to clinical training fellows? A retrospective study of the 20 years outcome of a Medical Research Council UK cohort'. *BMJ Open*, Volume 2:e001792.
- The National Centre for Post-Qualifying Social Work and Professional practice, (2018). CLAHRC East of England Fellowship Evaluation, Bournemouth: Bournemouth University.
- The Royal Society (2020). Diversity Data Report 2019, The Royal Society, London.
- SQW (2019). Knowledge Exchange Funding: a review of novel evaluation methodologies: final report, SQW/CityRedi.
- Ulrichsen, TC. (2015). Assessing the economic impacts of the Higher Education Innovation Fund: A mixed-method quantitative assessment, report for HEFCE, Higher Education Funding Council England (HEFCE).
- UKRI USAID (2018). Measuring Impact of Research and Innovation for Global Challenges, A report on the outcomes of a joint UKRI USAID workshop, 25-26 October, Panama.
- UKRI (2020). BBSRC Review of Fellowships Investments, s.l.: UKRI-BBSRC Skills and Careers Unit.
- UKRI (2020). Corporate Plan 2020-21, UK Research and Innovation.
- UKRI (2020b). The UK's research and innovation infrastructure: opportunities to grow our capacity, UK Research and Innovation.

UKRI (2020c). Diversity results for UKRI funding data 2014-015 to 2018-19, UK Research and Innovation.

UKRI (2020d). Research England Knowledge Exchange Framework: clustering and narrative templates, RE-P-2020-02, UK Research and Innovation.

Vitae (2011). Researcher development framework version 2, Careers Research and Advisory Centre (CRAC) Ltd.

Vitae (2015). Leadership Lens on the Vitae Researcher Development Framework 2012, Vitae, Cambridge.

Warry, P (2006). Increasing the economic impact of Research Councils, Advice to the Director General of Science and Innovation, DTI from the Research Council Economic Impact Group, 06/1678, Department for Business, Innovation and Skills.

Wellcome Trust (2020). Grant funding data report 2019/19, Wellcome Trust, London.

Zhitomirsky. V., McLean, M. and Klein, J. (2011). Patenting in the UK, PatAnalyse Ltd and Technologia Ltd.

## Annexe:

---

- A: Research landscape
- B: Mapping fellowship programmes
- C: Evaluation methodologies
- D: Learning from other evaluations
- E: Potential administrative data sources

## Appendix A: Research landscape

---

### The UK's research performance

There are various studies which examine the relative performance of the UK's research base and a range of measures or metrics are used, many of which are inter-related and can be impacted by wider factors. A key series is that undertaken by Elsevier with reports published in 2011, 2013 and 2016<sup>25</sup>, and most recently undertaken by BEIS (in 2019)<sup>26</sup>. These compare the UK to 7 other research-intensive countries (Canada, China, France, Germany, Italy, Japan and the UK) and to four fast growing nations (Brazil, India, Russia and South Korea). They break down the research base into funding and expenditure (inputs), numbers and mobility of researchers (human capital), numbers of journal articles published and citations received (research outputs), collaboration (research collaboration), ratio of research outputs to expenditure (productivity), and knowledge exchange.

These studies, undertaken before the UK left the EU and prior to the global pandemic, conclude the UK punches above its weight but it is losing ground over time to other countries as growth is slowing whilst in other countries it is increasing. The latest in the series of Elsevier and BIS/BEIS reports in 2019 found the UK accounts for: 2.5% of R&D funding, 4% of researchers, 6% of articles, and 14% of high-rated articles in the world; and the UK is ranked 3rd behind USA and China in terms of outputs. All indicators are down slightly from those reported in 2016: 2.7% expenditure, 4.1% researchers, 9.9% downloads, 10.7% citations, and 15.2% of highly-cited articles, so outcome measures are positive but falling slightly over time. These studies also conclude the UK research base is well-rounded and diverse with article outputs across all major research fields; and that the UK is key partner for global research collaboration and researcher mobility. However, the UK has a relatively low rate of global patents.

Most recently the impact of the global COVID-19 pandemic has brought research into sharp focus and has seen global collaboration and the UK research community at the heart of this.

### Importance of innovation to the UK economy

Over the last two decades there has been a growing realisation that the long run economic performance of nations, firms and industries depends on their ability to exploit technological innovation (Cohen, 2010). This has created a significant interest among policy makers in how policy can be designed to support innovation and encourage innovative firms to grow.

---

<sup>25</sup> <https://www.elsevier.com/connect/report-compares-uks-research-performance-with-key-nations>

<sup>26</sup> <https://www.gov.uk/government/publications/international-comparison-of-the-uk-research-base-2019>

Innovation can be defined as the first successful commercial exploitation of a new invention. The term covers both process of change and its outcome. Innovation processes are complex, uncertain, distributed and draw on a wide range of inputs to generate a wide range of direct and indirect outputs. They come in very different forms, with some drawing on formal research and R&D, while others relying on informal learning-by-doing and engagement with customers and suppliers. They can be positioned on a continuum from incremental to radical, and can generate new products, processes, services, or organisational structures. This complexity and heterogeneity makes innovation difficult to measure and research on innovation draws on a range of imperfect indicators to address the inadequacies of individual metrics (Hopkins and Siepel, 2013).

Research in the field of business level innovation has typically used an input and output measure of innovation to capture the subset of highly innovative firms. R&D spending was used as a measure that captured inputs to innovation, while the share of sales derived from new-to-market products was used as an output measure of innovation. As might be expected the upstream R&D measure was more closely associated with links to research, while the more downstream sales measure was more closely associated with links to suppliers and customers (Coad et al, 2014).

### **Role of graduates and postgraduates in the innovation process**

There is robust evidence on the wider importance of highly educated and talented employees in the innovation process. For example, the Coad et al (2014) study found that: (a) Highly Innovative Firms (HIF) have a significantly higher share of employment of science and engineering (STEM) graduates, and that this has a large positive influence on a range of performance metrics (more R&D, New-to-Market products, external co-operation and use of external information); (b) there is a strong persistence in the innovative status of firms, with HIFs remaining highly innovative; and (c) the growth process at the firm level starts with increased employment which leads to future increases in R&D spending and New-to-Market Products, and future increases in Sales. This causal chain suggests policy should avoid focusing exclusively downstream and consider what upstream capabilities need to be in place for increases in employment and ultimately sales to occur. The upstream focus will include employing and developing talented individuals and engaging them in the research process. This can occur directly, through the working in business or indirectly through the working in academic settings where new knowledge and innovation solutions are created, then diffused and commercialised throughout the wider ecosystem. This also addresses findings of HIFs research that they can be constrained by problems accessing managerial and technical skills, and financing (Couerduroy et al., 2012; Siepel et al., 2012; D'Este et al., 2012; Hutton and Nightingale, 2011).

The findings of the Coad et al (2014) report highlight that the value of investment in the research base comes primarily through the production of trained graduates and post-graduates who have the ability to solve complex technical problems and network more effectively. It is the production of 'talent not technology' that matters. Public investment in research generates talented graduates who leave the university system and go and work in industry. Their problem-solving skills reduce costs and increase economic benefits of innovation, increasing its demand and encouraging its exploitation and diffusion.



## Government focus and involvement

The government recognises the vital role of research and innovation and the need to maintain and build upon our strong position, and has a high-profile spending commitment for research. It has committed to doubling public R&D investment from £11.2bn to £22bn per year (2018 to 2024/25); and to spending 2.4% of GDP on research by 2027 (this figure includes private as well as public spending), and increasing to 3% afterwards. The government is looking to research to help solve the challenges of the 21st century including COVID-19, climate change, aging society, and levelling up across the UK.

This spending commitment as well as the plan for UK R&D has been most recently set out in the government's Roadmap (published in 2020). This sets out the vision for revitalising the UK's system of science, research and innovation that will strengthen the UK's global position in research, create long-lasting economic and societal benefits and help the UK to recover from the impacts of COVID-19. The 'mission' is to inspire and enable people from all backgrounds and experiences to engage and contribute to R&I, to nurture the whole 'system' of innovation, and to celebrate and showcase the UK's strengths to promote the UK as a destination for talent and investment. The roadmap asserts that bold changes are needed to ensure the system is fit for purpose and future proof. It focuses on several areas for action: raising research ambitions, driving up innovation and productivity, levelling up R&D across the UK, being at the forefront of global collaboration, developing world-leading infrastructure and institutions, and inspiring and enabling people and teams by attracting, training, and retaining diverse talent from all backgrounds. Each of these areas of focus has a series of ambitions. The focus on inspiring and enabling talented people and teams recognises that: careers in research can be lower paid and unstable; career pathways can be unclear; there is little support for movement between academia, industry and the public and third sectors or support for movement between countries; and the research culture in the UK needs attention particularly in terms of equality, diversity and inclusion but also collaboration, knowledge-sharing and collegiality. To support work on talent – the government is developing a new R&D People and Culture Strategy.

Growing the research talent pipeline and creating opportunities for diverse and flexible careers is a key part of the people focus within the roadmap, and essential to this is developing inspiring leaders who in turn can nurture and develop future talent. To support this aspect of the system the government aims to rethink the funding and wider support for postgraduate research, and to increase the support for early career researchers and to give them the skills, knowledge and experience needed to progress in careers inside or outside of academia and to increase the diversity within the research base. The UKRI Future Leaders Fellowships programme is an essential part of this ambition.

## Levelling up

Levelling up is a broad policy ambition based on the existence of geographic inequalities across the UK; and links to the emerging place-based focus for policy and research. London and the South East have higher economic performance than the rest of the UK, with some areas deemed to be 'left behind' and in need of 'levelling up' especially large towns and cities in former industrial regions, coastal areas and isolated rural areas. This

ambition has implications for research as the four factors targeted for levelling up are investment, transport, government and R&D.

It is widely acknowledged that R&D public spending is heavily concentrated in London and the South East (with also relatively higher concentrations in Scotland and the East of England) and so increased spending in other parts of the country could support levelling up. The issue for R&D activity and investment is that, although it can have benefits for the whole country, at least some benefits are concentrated in the area where it is carried out: direct spending by universities and institutes, and higher wages of staff; supporting knowledge-based local economy and greater local productivity; and generation of new knowledge. The government is therefore committed to take a greater account of place-based outcomes in how decisions are made about R&D, and is due to publish its R&D Place Strategy. However, some argue (for example the work of IFS, Davenport and Zaranko, 2020) that a fully equal allocation of R&D funding is not the best approach as not all places have the capacity to absorb funding and deliver top-end R&D.

### **Evaluation and research assessment**

The government want to 'spend their money wisely' so they are looking for better evaluation, and to link evaluation to funding in order to drive better impact. However, they do acknowledge that evaluation can drive negative behaviours, and there has perhaps been an overreliance on certain metrics or measures of success. It has been argued that there is a culture to pressure to publish and citations are ingrained in the reward system in HE, and this can distort the value of research and stifle creativity: publication can be regarded as an end in its own right and a justification for investment, and the way Research Excellence Framework (REF) has been implemented has established a risk averse/compliant culture. For example, 97% of the outputs to the REF in 2014/15 were text-based. This has led to the sector questioning if the way in which 'success' is measured is wrong and that different metrics and responsible research assessment approaches are needed, so that the focus can be placed on research quality. The next Research Excellence Framework is currently under way (REF 2020/21) but the Government is looking to evolve the REF and is launching a consultation.

### **Focus on research evaluation**

There is a move within the research system to demonstrate more clearly the impact of research. This is largely driven by the need for accountability as articulated in the Warry Report (2006) and has led to the inclusion of research impact into formal research assessment mechanisms such as the REF.

Work by Jones et al (2017) focused on the 2014 REF identified the drivers to research assessment as advocacy, accountability, analysis (what works), and allocation; and noted that these drivers can influence how impact is evidenced. In a recent study for Research England by RAND (Parks et al, 2019), six drivers for research assessment were identified including accountability which is the requirement to evidence that resources have been effectively and efficiently used. The others were: analysis, to understand whether research is effective; advocacy, to demonstrate the benefits of supporting research; allocation, to

determine distribution of funding; acclaim, to recognise the value of institutions and research, and create rankings; and adaption, to steer change. The authors note how adaption is a more recent driver for research assessment following an increasing focus on the wider societal impact of research. They assert that as the research landscape changes the reasons for assessing research at a national level will continue to develop and the emphasis or weight on each of the drivers will change.

The RAND report also highlights that:

- research outputs are likely to change with a greater diversity of output forms aimed at different audiences but that journal articles and conference contributions and authored chapters or books for those in arts and humanities are likely to remain the dominant forms of outputs in many disciplines. Researchers expect to produce more books, non-confidential research reports for external bodies (eg grey literature) and openly published peer reviews.
- Many factors can affect the volume and type of research outputs including: career stage and progression, personal preference, funder requirements, institutional incentives, desires to reach new audiences and create societal impact, and new technology. However it is widely recognised that the strongest influence on the type of outputs produced is discipline:

*‘There are significant differences in the forms of outputs being produced by researchers from different disciplines. For example, fewer arts and humanities researchers produce journal articles than researchers in other disciplines; while more researchers in the social sciences and arts and humanities produce book types: (i.e. chapters in books, authored books, book reviews and edited books), social media, blogs, podcasts and working papers. More researchers in medicine, health and life sciences, and physical sciences, engineering and mathematics, produce peer review, code, research datasets, and databases and preprints than researchers from the social sciences and arts and humanities. Some outputs are also highly specific to certain disciplines, for example analysis plans are mainly produced in the disciplines of medicine, health and life sciences, and software is mainly produced in physical sciences, engineering and mathematics. Although there are some differences between career stages, these are relatively minor compared to discipline-level differences. These differences are expected to continue in the future.’ (p7)*

- One of the key drivers for change in the research landscape is how research is assessed, but research assessment will remain dominated by peer review. However, there are challenges with peer review notably lack of equality and diversity, ways in which reviewers are rewarded, and cultural changes to recognise the value of open science. Suggestions for how research assessment should change going forwards include consideration of the diversity of outputs from academic research (beyond the outputs submitted to the REF which largely focuses on journal articles) and the needs of different disciplines.

The concern about the overreliance on publications in the assessment of scholarly research has led to the Declaration on Research Assessment (DORA) which emerged in the USA in 2012 and is now a worldwide initiative covering all disciplines. The Declaration

asserts the need to improve the ways in which research is evaluated to ensure scientific output is measured accurately and wisely and to recognise that outputs from scientific research are many and varied. Key to DORA is that Journal Impact Factor which is used frequently as a primary measure to compare scientific outputs was not intended for that purpose and is not a suitable tool for research assessment. Criticisms include citations are highly skewed, it is field-specific, it can be gamed by editorial policy, and the data used are not transparent or open to the public. It is acknowledged that peer reviewed research papers will remain a key research output informing research assessment but recommends research is assessed on its own merits not on the basis of the journal it is published in, that journal-based metrics are not used to assess individual scientists' contributions, that the value and impact of all research outputs are considered including datasets and software, and a broad range of impact measures including qualitative indicators such as influence on policy and practice are used. DORA also calls for the representation of researchers in the design of research assessment practice to reduce inequalities.

## UKRI mission and evaluation stance

The UKRI was launched in 2018 to build on the strengths of the nine research councils and enable more ambitious and interdisciplinary research and innovation. It is the largest public funder of R&D in the UK and is central to the delivery of the UK's R&D Roadmap. A key priority is to connect the breadth and depth across the R&I system, getting disciplines and sectors to work closely together. The UKRI's Corporate Plan for 2020/21 (UKRI, 2020) sets out its vision for the R&I system, its mission in stewarding that system, and areas of focus (and related goals and ambitions).

- Vision: *'for an outstanding research and innovation system in the UK that gives everyone to contribute and benefit, enriching lives locally, nationally and internationally'... to shape a dynamic, diverse and inclusive system.*
- Mission: *'to convene, catalyse and invest in close collaboration with others to build a thriving, inclusive research and innovation system that connects discovery to prosperity and public good'.*
- Four areas: convene and catalyse, incentivise, conduct, and invest. The invest area is described as *'we will invest in people, ideas and infrastructure, through a portfolio of investments that ensures public benefit from the system as a whole, informed by engagement and evidence. We will fund well – efficiently and effectively'.* This includes a priority objective to invest in the best people, ideas and infrastructure. Beyond 20/21 UKRI ambitions are *'support the development of the positive, diverse, interdisciplinary research and innovation culture that is central to the success of our system and will back our researchers and innovators to do their best work, delivered with minimal red tape and supported by a high-quality research environment'.*

The corporate plan notes that as part of the commitment to transparency and accountability UKRI is developing a performance and impact framework so it can judge success 'robustly and transparently' whilst also managing its performance. The framework will include outcomes and impacts. UKRI are also developing a People and Culture Strategy (working with BEIS). The corporate plan also outlines the role of UKRI in addressing shared

priorities across the nine research councils, to tackle challenges that require collaboration across disciplines through cross-cutting funds. The FLF is part of UKRI cross-cutting funds (along with Global Challenges Research Fund, Newton Fund, Industrial Strategy Challenge Fund, Strategic Priorities Fund).

The plan provides details of the planned performance and impact framework (to be developed in 21/22 onwards) and how this will consider not just inputs and activities but also outputs, outcomes and wider impacts, noting challenges such as the time taken for impacts to emerge and difficulties in attributing reliably to specific interventions, particularly when they are diverse, novel and unpredictable:

*‘This means outcomes and wider impacts are not best tracked through a set of pre-determined performance metrics, and there are potential adverse effects of becoming overly reliant on proxy metrics to judge the value of research and innovation on our community, our culture and our effectiveness. We should be cautious to avoid creating unnecessary reporting requirements and incentivising the wrong behaviour’ (p50).*

At present UKRI take a pragmatic approach to addressing in-year performance using a balanced scorecard approach assessing against the UKRI’s key objectives. This involves making balanced nuanced assessments using a range of leading and lagging performance measures which recognise the differential nature of impact across the disciplines, and drawing on quantitative and qualitative evidence including corporate performance data, programme monitoring data, outcome monitoring data, analysis of internal datasets, collecting stakeholder views and opinions.

In addition, each research council has been developing their approaches to evaluation of the research and activity it funds, and around possible impacts particularly economic and societal impacts. For example, the ESRC had a programme to explore and measure the impact of social research and to understand how impact is generated in recognition of the broad, diffuse and non-linear impact of social science and that impacts can be indirect, hard to quantify and take time to emerge. This programme generated a number of reports: Taking Stock (2009), Branching out (2011), and Cultivating Connections (2013). The programme highlights the importance of underpinning theoretical frameworks that recognise the complexities and multifaceted nature of social science impact, that multi-dimensional approaches are needed to give comprehensive accounts of the impact of research, data and people, and that the processes through which impacts are generated and the context also need to be taken into account. It identifies three categories of impact: instrumental, influence on development of policy, practice, service provision, law or behaviour; conceptual, contributing to understanding; and capacity building, on technical or personal skill development.

## Challenges for research in the UK

Stakeholders assert that the UK research faces a number of challenges which are likely to be heightened by the pandemic, many of these relate to funding – difficulties securing funding and particularly sustainable funding (multi-year funding settlements to enable research to tackle large-scale national challenges):



- Importance of the private sector in research and innovation. The private sector accounts for over half of all R&D spend and just under half of all funding. For example, the 2016 Elsevier report notes 65% of R&D expenditure was in the business enterprise sector (with 26% in the higher education sector and 7% in government sector), and 48% of funding input from business and enterprise (28% government, and the rest is investment from overseas including EU grants, and non-profit sector). There are concerns about how the economic effects of pandemic will impact on funders, researchers and UK HEIs, and particularly on the private sector.
- There are concerns about government research funding during economic retraction, and whether government can deliver its 2.4% spend commitment. There is an expectation in the sector that Treasury will want to reduce spending; and there will be other priorities for government (other than higher education). There are fears that the spending commitment is at risk due to economic, political and public pressure. The spending commitment was set out in the November Spending Review (in 2020), in the UK R&D Roadmap and the recent Integrated Review, plus in February 2021 the government announced their intention to launch the new £800m independent Advanced Research and Invention Agency (ARIA) which will fund high risk, high-reward research<sup>27</sup>, and in April 2021 it announced £250m to support research<sup>28</sup>. However, the Spring 2021 budget announced significant cuts to the research budget, particularly to the Overseas Development Assistance which is a key funder for international collaborative research<sup>29</sup>.
- The importance of cross-subsidy (such as the income from fees from international students) in research funding within higher education, and how this might be affected by the pandemic and its economic repercussions and by the UK's exit from the EU.
- How best to allocate resources in a fair and inclusive way and which involves measuring impact in a fair and inclusive way.

Other challenges recorded in the literature include:

- inequalities in researcher pool, with women underrepresented and falling behind in the pandemic and this impacts on careers, and UKRI diversity data (UKRI, 2020c) shows differences in the proportion of applicants, differences in the award rates, and differences in the award amounts by gender and ethnicity.
- risk aversion,
- limited strategic focus,
- lack of interdisciplinarity,
- bureaucracy.

<sup>27</sup> <https://www.gov.uk/government/news/uk-to-launch-new-research-agency-to-support-high-risk-high-reward-science>, accessed 23.02.2022

<sup>28</sup> <https://www.gov.uk/government/news/250-million-additional-funding-to-boost-collaboration-and-protect-ongoing-research>, accessed 23.02.2022

<sup>29</sup> <https://www.ukri.org/news/ukri-required-to-review-official-development-assistance-funding/>, accessed 21.02.2022

There are moves to try to reduce bureaucracy in the system and a recent policy paper (BEIS/DfE, 2020) sets out measures to remove unnecessary bureaucracy and the need for a system wide review of the causes of unnecessary bureaucracy. Measures include changes to selection processes for research funding involving simplifying eligibility criteria, streamlining the number of grant schemes, streamlining the application process for grants, and moving to a single approach to setting out track record. Also measures include changes to capturing research outcomes involving harmonising reporting requirements across UKRI, reducing the number of questions for mandatory reporting for Researchfish, and reviewing the approach to outcomes monitoring to be minimally demanding.

## Appendix B: Mapping fellowship programmes

---

To understand the large and diverse landscape of fellowships a desk review was undertaken to identify programmes, schemes and initiatives aimed at developing research capability in the UK, supporting ECRs, and developing research leaders. This review sought to understand the number, scope, nature and goals of these other programmes, as they have the potential to impact upon the same areas as those outlined in the goals and ambitions for the FLF programme (and developed in the Theory of Change for the FLF programme). This is important to understanding the feasibility of attributing impact to FLF which is a key challenge for the full evaluation.

A secondary aim of the review of fellowships and innovator programmes was to compare key aspects of the FLF programme with these other programmes and to potentially identify those that could provide an appropriate benchmark or comparison group for the impact evaluation. Stakeholders agree that the FLF programme has a number of characteristics which make it relatively unique in this field and thus challenging to find contemporary programmes:

- Large size in terms of both funding amount and numbers of awards – the total funding for the programme is £900 million and the target number of fellows is 550. Each individual fellow is generally able to apply for up to £1.5m (although there is no limit to the amount that can be requested) in a personal award to support their research and develop their career. Round 6 guidance notes that the programme has funded fellowships from £300,000 to over £2m and that there is no preference (in assessing applications) for lower or higher costed proposals. Across rounds 1 to 4, 298 Fellowships have already been awarded. Round 5 is currently being assessed, and Round 6 (final round) is open for applications.
- Long period of funding – the funding and wider support is available for up to seven years. The funding is provided over four years initially with the possibility of extending funding and support for a further three years after a further assessment exercise.
- Wide(r) diversity of host organisations - fellows can be located in commercial businesses as well as academic organisations (including Higher Education institutions, independent research organisation (IRO), Catapults, and Institutes). The funded work of fellows could cover innovation, incubator activity as well as research. To be successful, fellows must demonstrate that they have significant support from their hosting organisation (this commitment is part of the assessment process) which will include dedicated time on their FLF activities and thus reduced requirements for teaching or other business activities and a commitment to offering an open-ended position at the end of the fellowship. All hosts must be UK-based.
- Not limited to specific disciplines – the programme is intended to support individuals in all areas of UKRI's remit including multidisciplinary and interdisciplinary projects.



Individuals can apply from any discipline and the research and innovation funded can cross disciplines.

- Wider eligibility criteria – eligibility for funding is wider than found in many early career researcher programmes, as applicants are not required to hold a PhD (to reflect the wider host organisations the programme aims to involve) nor have achieved a PhD within a set period or to hold a permanent or open-ended position. Applicants can be UK-nationals or come from abroad (with international applicants eligible for a Global Talent visa under the exceptional promise category). Additionally, the programme itself is designed to be flexible in order to attract a wider pool of applicants including those wishing to work part-time.

A third aim of the review was to explore the extent to which these programmes have been monitored and/or evaluated and identify the methods, measures and data sources used and any challenges encountered to see what can be learned for an evaluation of the FLF programme.

## Approach

A set of criteria was established to identify relevant programmes and a number of key sources were searched including websites of known key funders, key HE sector bodies, organisations dedicated to supporting researchers. The UKRI FLF programme team also provided details of a mapping exercise they had conducted.

The resulting map of programmes was not intended to be exhaustive and would not have captured all current and recent opportunities particularly smaller/less well-known and irregular opportunities. It was intended to give an indication of the support available to potential research leaders and innovators. It is also likely to be heavily skewed towards more ‘academic’ programmes hosted in academic settings. The mapping process was somewhat challenging as programmes may not take place on a regular basis, may have been paused for some time, or can change their names and/or be replaced over time. Thus some programmes in the map may no longer be accepting applications, but may still be running (supporting recipients) and have been the focus of evaluation activity.

## Mapping insights

The mapping exercise indicates that there are many opportunities for funding and -support for researchers, particularly those in their early careers. Overall, between 150 and 170 programmes were identified and mapped, of these 120 were specifically aimed at early career researchers. The rest either aimed at all career stages or those more established (mid-career, or independent researcher). Each individual funder often offers an array of different schemes each year with varying aims, eligibility and support provided.

## Digging deeper

There appear to be five categories of funders. However, as some funds/programmes are jointly funded by several organisations, these can cross categories and can be deliberately designed to do so. For example, the APEX – the Academies in Partnership in Supporting

Excellence in Cross-disciplinary - research award scheme brings together various funding bodies. Similarly, many large trusts and charities which fund support for early-career researchers have linked academic or industry partners. The categories are:

1. UK Research & Innovation (UKRI), and its seven Research Councils, Innovate UK and Research England, and the various Catapult centres (established by Innovate UK)

The research councils offer a range of fellowship schemes, many of which are aimed at early career researchers within specific fields. They tend to cover salaries of researchers for a number of years and require individuals to have a PhD, but can provide support in the range of £250 to £500k. There are some fellowships providing larger amounts of funding such as the David Phillips Fellows (BBSRC, currently paused) providing £1 million in support over 5 years, Career Development Awards (MRC) providing up to £1 million over 5 years, and Senior Clinical Fellowship and Senior Non-Clinical Fellowships (MRC) providing £2 million over 5 years. These however tend to fund a small number of individuals and given the disciplinary nature of the funders are more narrowly focused than FLF.

The new Innovation Scholars Secondment scheme has some degree of similarity with FLF in that it is focused on porosity, career development and knowledge exchange and a PhD is not a pre-requisite. However it is much smaller in value and duration of support (up to £300,000 for 3 years), is open to those at all career stages, and is essentially a secondment scheme so operates very differently. It also appears to be narrowly focused (biomedical sciences). Within this category there are also some industry/enterprise focused fellowships such as the STFC Innovations Partnership Scheme Fellowships and Enterprise Fellowships, and Ser Cymru's Future Generations Industrial Fellowships which perhaps start to capture some of FLF's ambitions for porosity with industry and knowledge exchange. But these are much smaller in scale, not necessarily focused on early career researchers and again narrowly (disciplinary) focused.

Perhaps the most similar to FLF are the European Research Council Starting Grants (or Starting Independent Researcher Grants). These Starting Grants are aimed at early career researchers who are ready to work independently, to provide long-term funding to support excellent investigators and their teams to pursue ground-breaking, high quality/high risk research which is expected to lead to advances at the frontiers of knowledge. These are similar in scale to the FLF programme, providing up to €1.5m (plus a further €1m to cover start-up costs) for a period of 5 years. They are open to researchers of all nationalities, to a wide range of disciplines covering physical sciences, life sciences and social sciences and humanities, and multi/interdisciplinary proposals are encouraged; but a PhD is a prerequisite (and applications must be made within 2 to 7 years of completion). Also, although hosts can be private organisations this doesn't appear to extend beyond private laboratories; and whilst more than 400 awards are made each year just 62 grants in 2020 were to UK hosts. They are highly competitive, and in 2020 there were 3,272 applications for 436 awards (at a cost of €677m), giving a

success rate of 13 per cent<sup>30</sup>. The names and host institutions of the successful researchers are posted on the ERC website.

This group of ERC Starting Grant Fellows could form a potential comparison group for the impact evaluation for FLF programme fellows operating in academia, although there are still some differences between the two programmes.

2. Trusts, Foundations and charities many of which are health focused. These include Alzheimers Research UK, British Heart Foundation, Cancer Research UK, Diabetes UK, Guarantors of Brain, Marie Skłodowska-Curie, Versus Arthritis, and Wellcome Trust; but others include Leverhulme Trust, NESTA, The Royal Commission for the Exhibition of 1851, Biometrika Trust, Cambridge Philosophical Society, Daphne Jackson Trust, and the Wolfson Foundation.

These fellowships tend to be quite narrowly focused to a specific field/discipline to align with interests of the funder. Several of these schemes offer support of around £25,000 to £50,000 but support can be in the region of £200,000 to £500,000 or even up to £1m in the case of Wellcome Trust Investigator Awards but these are aimed at more established researchers. Some schemes don't provide notional or maximum costs but note they will cover the salary costs of the fellow plus other costs and expenses. For some schemes this can involve additional staff such as postdoctoral researchers, PhD students, research assistants and technicians (eg Cancer Research UK and British Heart Foundation Fellowships) to start to form a team around the fellow. Generally, the numbers involved in each programme are small or not given.

The exception is Leverhulme Trust Early Career Fellowships programme, where over 100 Fellowships are awarded each year and cover all subjects with the exception of clinical practice, and are aimed early career researchers who will be resident in the UK for the period of the Fellowship. Within this funding category they appear to be the most similar in volume and eligibility to FLF. However, as they intend to support building an academic career in the UK and provide £100,000 of funding and support for 3 years, these Leverhulme Trust fellowships are narrower in scope than FLF which supports research and careers beyond academia, and involve considerably less financial support than offered in FLF. The Leverhulme Trust has larger awards of £1m for Research Leadership Awards but only a couple of awards are made each year.

Also of interest are the Royal Commission for the Exhibition of 1851 Industrial Fellowships which although the support involved is modest (£25,000 per year for 3 years) and the volume is small (10 awards per year) do have a focus on industry. The aim of this fellowship programme is to encourage profitable innovation and creativity in British Industry to the mutual benefit of the Fellow and the sponsoring company. However, projects are limited to those in a science or engineering discipline.

3. Learned societies. These include Academy of Medical Sciences, British Academy, British Pharmacological Society, Royal Academy of Engineering, Royal Geographical

---

<sup>30</sup> [https://erc.europa.eu/sites/default/files/document/file/erc\\_2020\\_stg\\_statistics.pdf](https://erc.europa.eu/sites/default/files/document/file/erc_2020_stg_statistics.pdf)

Society, Royal Astronomical Society, Royal Society of Chemistry, The Royal Society, and Royal Society of Edinburgh

The financial support provided through learned society fellowship programmes tends to be small, for example less than £10,000 to cover some costs or replacement salary, can be very narrowly focused, and with a small number of awards made each year. Some of the learned societies offering larger numbers of fellowships or equivalent include the Academy of Medical Sciences who have offered almost 500 Starter Grants for Clinical Lecturers since 2008 to enable research active Clinical Lecturers to pursue research by providing modest research funds of up to £30,000 for up to 2 years. Other larger volume programmes include The British Academy Postdoctoral Fellowships (over 30 a year), the Royal Academy of Engineering Sainsbury Management Fellowships (375), and The Royal Society University Research Fellowships (30 to 40 a year).

More generous fellowships in this category include the Royal Academy of Engineering Fellowships offering up to £500,000 across 5 years, Wellcome Trust's Sir Henry Wellcome Postdoctoral Fellowships offering up to £300,000 across 4 years<sup>31</sup>, and Sir Henry Dale Fellowships also offered by Wellcome Trust and The Royal Society offering up to £1.2m to cover salaries and research expenses, for 5 years with a potential extension for a further 3 years. The Sir Henry Dale Fellowships have a similar volume, and level of support to that offered by FLF. They too are aimed at early career researchers (or as the Wellcome Trust note 'early-independent or intermediate researchers') to build their own independent research programme and team and establish themselves as an independent research leader in their field. However these awards are restricted to PhD holders with significant postdoctoral research experience (no limits on time since PhD award), who are expected to have already made significant research contributions such as publications, patents, software development, and are starting to lead their own research. They are also limited to health related scientific research enquiry.

Sir Henry Dale Fellows could potentially provide an alternative control group for FLF, particularly for those in academia and STEM disciplines, despite the programme closing to new applicants after 2021. The names and hosts of fellows (and year awarded) are published on the Wellcome Trust website.

There are some fellowships funded by learned societies with specific ambitions: aimed at industry including Royal Academy of Engineering Industrial Fellowships, and Royal Society Industry Fellowships; aimed at leadership development (Health Foundation and Academy of Medical Sciences' Clinical Scientist Fellowships); and the Apex awards aimed at supporting excellence in cross disciplinary research. However, no programme attempts to cover all these aspects within one scheme.

4. Universities (individually and in groups) and Research Institutes. These include the Consortium of Welsh Universities, the GW4 Alliance (a collaboration between the

---

<sup>31</sup> Sir Henry Wellcome Postdoctoral Fellowship and Sir Henry Dale schemes will close to new applications in 2021, with March 2021 as the final round/cohort.

Universities of Bath, Bristol, Cardiff and Exeter), and the European Molecular Biology Laboratory – European Bioinformatics Institute, Faraday Institution, and (Wellcome) Sanger Institute. University fellowships are often funded through alumni bequests/endowments, or are badged as Presidential, Chancellor or Vice-Chancellor's Fellowships or as 'Crucible' Fellowships.

These tend to offer a salary or Stipend ranging from £25,000 to £60,000 and some offer additional expenses to cover travel or other research costs. They generally offer support for between 3 and 5 years and offer a handful of fellowships each year. Larger schemes are the Anne McLaren Research Fellowship and Nottingham Research Fellowships at the University of Nottingham with 100 fellowships.

#### 5. Health sector including the Academic Health Science Networks, Health Education England, Health Foundation, National Institute Health Research, and NHS.

In addition to the health focused programmes offered by the relevant research councils, these bodies also offer fellowships. These programmes tend to cover salary and research costs but often don't give a clear indication of budget. Support is generally provided for 3 to 4 years and several are aimed at any career stage (including early careers). There are a number of interesting schemes with very specific aims such as developing leaderships skills (eg Darzi Fellowships), promoting researcher mobility (Marie Skłodowska-Curie Early-Stage Researcher Fellowship), developing research skills in clinicians (NIHR Academic Clinical Fellowships), and spreading/scaling and implementation of innovation (eg Accelerator programmes such as NHS Innovation Accelerator and DigitalHealth.London's Accelerator). Again no one programme attempts to cover all aspects.

There are also a few programmes and funders which do not fit neatly into the above categories. The funders and programmes include BEIS and the programmes it directly manages such as the Newton Fund (which is delivered through research councils and learned societies, and builds partnerships between UK researchers and those in countries to take on sustainable development priorities), also the Rutherford Fund (to attract early career and senior researchers from developing countries to the UK) and Knowledge Transfer Partnership (a programme partnering academics and researchers with businesses to improve competitiveness and productivity, managed by InnovateUK); plus the Government Office for Science, Society of Chemical Industry, and UK Space Agency and the development programmes they offer.

## Evaluation activity

Very few of the identified (and mapped) programmes - around 20 - appear to have been evaluated or state that some form of monitoring and/or evaluation is planned or is underway. Where evaluation has or is planned to take place this has tended to be programmes funded by UKRI or its constituent bodies, or those funded by health bodies.

Generally, it is difficult to find evidence of published evaluations. In some cases, programmes have been monitored and reports are available; and, in others, programme guidance can explicitly state that data will be collected and held for monitoring and

evaluation purposes. In these cases, no full monitoring or evaluation reports appear to be publicly available but monitoring details are published in the annual statements or reports of the funders. The monitoring has largely been limited to input level data such as numbers of awards, sums spent and, in some cases the characteristics of recipients and case studies of individual projects or recipients/fellows/ prize winners. This monitoring can demonstrate to trustees/key stakeholders how the funds have been spent but there is perhaps an unwritten expectation that the funder will indirectly benefit from this spending on research(ers) and innovation(ers) with no expressed causal link and no evidence of impact. In most cases when exploring the nature of the programme (aims, eligibility, support etc) there is no mention of evaluation.

In some cases, there has been a process evaluation undertaken to understand the nature of the programme and particularly the support received, in order to make improvements for the future. In a small number of cases, there has been an attempt to undertake an evaluation of the impact of the programme but this has largely been limited to surveys of fellows and individual case studies rather than a full counterfactual impact study.

The lack of evaluation of fellowship programmes could be due to the complex nature of funding arrangements (particularly where there are joint or multiple funders), the scale of programmes (which can be quite small or very large), or the varied nature of the activities funded. However there appears to be a move towards impact evaluation of programmes, with some (particularly larger, UKRI and research council funded) programmes noting upcoming monitoring and evaluation.

## How does FLF fit into the landscape

Comparing the programmes found in the mapping process it is evident that the size of the Future Leaders Fellowships programme is much larger than many other programmes. The anticipated total number of FLF fellows is 550; whereas many other fellowship schemes especially those not UKRI funded have less than 20 fellows per year (usually 1 to 5). The exception to this is the longstanding Early Career Fellowships scheme funded by the Leverhulme Trust which has had well over 1,500 recipients, around 150 each year. This appears not have been evaluated.

The lengthy period of funding (over a seven-year period) provided by FLF is also quite unusual compared to other programmes. Most range from 2 to 5 years and are often linked with the academic year. A few Cancer Research fellowships can last 6 years, whilst some senior chair positions can last up to 10 years.

Many of the programmes identified are based solely in academia. However, the Future Leaders Fellowships programme is not completely unique in having host institutions both inside and outside of academia, with some not-for-profit institutions hosting academic funded recipients and other programmes funding roles based in several locations (including in clinical settings). The multi-disciplinary nature of the Future Leaders Fellowships programme is also relatively distinctive as is the strong cohort focus. It is common that other programmes are unique to the disciplines/industries associated with the funder.



## Appendix C: Evaluation methodologies

### Potential evaluation approaches

The review of evaluation approaches indicates that there are different ways that evaluation methods can be grouped. The groups are not always mutually exclusive and sometimes there is a need to use multiple approaches or synthesise methods together. The grouping depends on several dimensions including purpose, feasibility, and method.

### Key evaluation categories

Once a programme, or part of a programme, is identified for evaluation, the next step is to identify the type of evaluation to conduct. These include predominantly process, impact and value for money evaluations, and also formative and outcome evaluations. Key sources which discuss evaluation approaches are the Magenta Book (HM Treasury)<sup>32</sup>, CDC Programme Evaluation<sup>33</sup> and Humans of Data<sup>34</sup>.

- Process evaluation monitors programme activity and it can take place during or as soon as the implementation stage of a programme starts. The aim of a process evaluation is to determine whether programme activities have been applied correctly and as intended. It makes it possible to review activities and outputs, track programme progress, and pickup on early warnings of any issues that may occur. This information can be used to finetune the logic model or Theory of Change in future implementations. For example, a process evaluation can answer if project activities reached target groups, if they are being implemented as intended and if any changes have been made to the intended activities.
- Impact evaluation takes place during the runtime of a programme in certain periods or at the end of a programme and beyond. It helps assess if, and to what degree, the programme meets its goal and how well have the desired short-term (or long-term) changes been achieved.
- Value for money evaluation may inform whether the intervention was indeed a good use of resources given the benefits it produces and its costs. There are two levels in such an evaluation. Firstly, it compares the costs and benefits of the programme to see if the benefits are larger than the costs. Secondly, if the programme's benefits are deemed to

<sup>32</sup>

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/879438/MT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/MT_Magenta_Book.pdf)

<sup>33</sup> <https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf>

<sup>34</sup> <https://humansofdata.atlan.com/2017/04/7-types-of-evaluation/>



outweigh the costs, it compares the cost effectiveness of this programme to the cost effectiveness of alternative programmes.

- Formative evaluation is used when a new programme is being developed or when an element of an existing programme is being altered (e.g., used on a different population or in a new setting). Its purpose is to help identify whether the new programme or new programme element is feasible and/or necessary, and whether an evaluation is possible based on the nature, goals, and objectives of the programme. It is an important tool because it sheds light on whether any changes to the programme need to be made before it starts, which maximises the likelihood of its success.
- Outcome evaluation takes place after the programme has contacted at least one subject in the targeted population. Its aim is to measure whether, and to what extent, the programme affects the population when it comes to a certain outcome. Therefore, it allows us to tell whether the programme is effective in meeting its objective. For example, it can help pinpoint factors outside the programme that may have intervened in the desired change and, conversely, if the programme has caused any unintended changes.

## Impact Evaluations: Theory based vs experimental approaches

Looking more closely at impact evaluations, a key ambition for the full evaluation of the FLF programme, there are many different approaches that can be adopted but these tend to fall into either theory-based or experimental approaches. According to the Magenta Book (HM Treasury)<sup>35</sup>, the choice between theory-based methods and experimental methods depends foremost on whether it is possible to compare groups affected and not affected by the intervention.

### Theory-based methods

Theory-based methods are used to explore net impacts by discerning the causal links that are likely to bring about change by an intervention. Unlike experimental methods, theory-based methods are unable to give a numeric estimate of the effect size of the intervention. However, they can still be qualitative and quantitative. Theory-based evaluations answer to what extent did change take place and why did it happen, taking into account the context of the intervention being administered. That is why they are particularly well suited for the evaluation of complex interventions or simple interventions in complex environments. These methods can verify that an intervention had the desired effect in a context when measuring the magnitude of the effect is difficult.

In that sense, most methods that belong to this group do not aim to give conclusive evidence that a change can be entirely attributed to an intervention. Instead, they aim to provide evidence for whether the intervention contributed to a measurable change, explain why it was successful, and consequently advise on how that change can be translated to

<sup>35</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/879438/HMT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/HMT_Magenta_Book.pdf)

other contexts (e.g. different population, place, or time period). Theory-based methods often require experts to help decide on the appropriate approach.

Some of the most highly regarded theory-based methods include: realist evaluations, qualitative comparative analysis, contribution analysis, process tracing and simulation modelling.

#### ■ Qualitative comparative analysis:

Qualitative Comparative Analysis (QCA) combines both qualitative and quantitative methods to gather in depth insight into different cases, capturing their complexity whilst at the same time attempting to produce some form of generalisation. It involves systematically comparing cases as a set of conditions and outcomes in order to find patterns as well as to identify redundant or idiosyncratic conditions. This results in the identification of explanatory patterns of success and failure and informs future cases accordingly.

*“QCA is often portrayed as a bridge builder between contextualization in terms of accounting for the idiosyncratic nature of specific cases and generalization in terms of unravelling trends across these specific cases” Pattyn et al (2017, p.58) <sup>36</sup>*

QCA can be used when the number of cases is too small for quantitative analysis but too large to do in depth qualitative methods. Instead of focusing on averages, QCA allows for contextual explanation and causal complexity. By relying on set theory (sets are groups in which the cases are members), QCA can identify conditions that are sufficient and/or necessary for a certain outcome to occur - that way QCA can disentangle relations between contextual conditions and outcomes. The outcome in QCA can either be binary, a sliding scale, or multi-value, so it is able to account for both qualitative and quantitative variations.

The benefits of QCA include transparency and replicability, increasing internal validation. It is also more about learning rather than accountability i.e., better at answering why something works (or doesn't) rather than answering what works in the first place. It is also able to handle a substantial degree of variation between cases (no need for everything to be 'held constant') as long as there is a degree of comparability. The “learning” attitude towards unsuccessful cases may be challenged as some may not be willing to share information about failures. Since failure constitutes an important part of QCA, it needs to be analysed even when stakeholders want it to be overlooked<sup>37</sup>.

There are several shortcomings to the QCA; for one, it does not open the ‘black box’ of causality, as it is only able to show why a combination of factors leads to a certain outcome, making it open to interpretation. It is therefore better at capturing correlation than causation. It also is also static with no time dimension. Practically, it requires the

<sup>36</sup> <https://journals.sagepub.com/doi/pdf/10.1177/1098214017710502>

<sup>37</sup> *ibid*

collection of comparable data and when the outcome of interest is subjective, it requires value judgement to label them on a scale from zero to one<sup>38</sup>.

There are conditions that allow QCA to be more effective. It is best used when there is a 'correct' balance of similarities and differences between the cases, the former to make them comparable and the latter to capture variation. To avoid any issues, it is best that the survey design is done as early as possible and that the surveys include extensive information, even if some do not end up being used. Finally, it is important to consider causation paths for each case after the results are out<sup>39</sup>.

#### ■ Realist impact evaluation

Realist impact evaluation is a theory-based evaluation approach that starts with clarifying how a programme is understood to cause/contribute to outcomes and impacts. It is different from other methods of impact evaluation because it has 'realist' assumptions: not everything works for everyone everywhere and all the time as context matters and contributes to differences in outcomes. As a result, it is important to understand what works and what does not work and why a specific context facilitates something to work but not another. The question asked is, therefore, how and why does this work, for whom, to what extent, in what respect, in which circumstances, and over what duration?<sup>40</sup>

Realist evaluation or Critical Realist approach is really a suite of methods and Theory of Change sits within these. It seeks to set out from intervention design the intended mechanisms through which interventions work. The approach focuses on mapping the connections between activities and outcomes within an intervention, to generate hypotheses about how the intervention will achieve the desired change. It therefore makes explicit the underlying assumptions about how programmes or interventions are expected to work (Rogers et al., 2000) and thus enables a more systematic focus on explaining how and why an intervention works (or does not work). This then provides a benchmark against which to test interventions to understand if the theory held true in practice and what factors in the context affected what is seen.

As this is an impact evaluation method, it focuses heavily on causation and attribution. Rather than comparing changes for treatment and control groups, the realist evaluation compares whether a programme works differently in different localities or for different population groups and why that is the case. Because realist impact evaluation is more a way of thinking than it is a method, it can be incorporated within any impact evaluation. Some conditions make it easier to use, including when the evaluation is for a new initiative or a pilot programme that produces the desirable outcome, but it is not clear how or for whom. It is also useful if there is a need to understand how things will change when the programme is taken to another context. It works well when a programme shows mixed patterns of outcomes as it can help explain why these differences occur. It

<sup>38</sup> Ibid

<sup>39</sup> Ibid

<sup>40</sup> <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>

can take on both a qualitative and quantitative approach – and can be used on small or large data<sup>41</sup>.

Realist impact evaluation is not appropriate, for example, when stakeholders are more interested in *whether* a programme works rather than in *how* the programme works. It is also not appropriate when the purpose of the evaluation is to discern the net effect and if the programme is implemented in a simple setting. This method is best used when there is a clear purpose for why the question needs to be answered and when the answer will be used for something specific (policy or practice). Its advantage over experimental designs is that it can be used in cases when the intervention is small scale (small sample size) and when the intervention is universal (there is no control group). It has the most added value in certain circumstances including when there is a clear initial programme theory, when there is a longer-term objective that allows for iterations, and when the programme is implemented across multiple sites or multiple groups<sup>42</sup>.

#### ■ Contribution analysis

Contribution analysis explores causal effects through assessing the contribution a programme is making to an observed result – it tests the theory of change, and takes into account other influences. This method is useful when the aim is to disentangle cause and effect when experiment is not an option. It provides reasonable evidence on the contribution of the programme and has to be motivated by a sound theory of change. Kotvojs & Shrimpton (2007, page 1) describe using contribution analysis *"as a means to consider progress towards outputs and intermediate and end outcomes."*<sup>43</sup>

The uniqueness of contribution analysis is that it is designed to minimise uncertainty about the contribution the intervention is making to the outcome. This is done through increasing the understanding of why the outcome has or has not occurred, taking into account the impact of the intervention separately from other internal and external influences. Contribution analysis is, therefore, better at answering questions such as: Has the intervention influenced the observed result? Has the intervention or programme made an important contribution to the observed result? Why did the result occur? What was the contribution of the intervention? It can therefore shed light on whether it is reasonable to infer that the programme resulted in a change, how well the evidence suggests that the change has been made, and the conditions needed for this intervention to succeed<sup>44</sup>.

The benefits of contribution analysis include providing evidence for causal relationships outside of experimentation. It can also be used as an iterative process, which permits the mitigation of unforeseen circumstances. However, it is better at capturing direct influences over indirect influences and might require revising the theory of change as result of the iterative process. It is most effective when a strong, appropriate theory of

<sup>41</sup> [https://www.betterevaluation.org/en/approach/realist\\_evaluation](https://www.betterevaluation.org/en/approach/realist_evaluation)

<sup>42</sup> Ibid & <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>

<sup>43</sup> [https://usaidlearninglab.org/sites/default/files/resource/files/mod17\\_ausaid\\_fiji\\_case\\_article.pdf](https://usaidlearninglab.org/sites/default/files/resource/files/mod17_ausaid_fiji_case_article.pdf)

<sup>44</sup> [https://www.betterevaluation.org/en/plan/approach/contribution\\_analysis](https://www.betterevaluation.org/en/plan/approach/contribution_analysis)

change is developed, especially with regards to determining the relevant cause - effect question in any specific context and when other factors that may influence outcomes outside the programme are identifiable.<sup>45</sup>

Contribution analysis can shed light on the validity of the theory of change, either by reaffirming it or by requiring it to be revised. However, it is not intended to uncover and display a previously implicit or 'hidden' theory of change. Indeed, it works best when the theory of change is clearly defined and when the implementation is uniform, with no scope for variation. The results provided by contribution analysis are not absolute proof; they simply provide a line of logic or reasoning from which plausible results can be concluded with some level of confidence.<sup>46</sup>

## Experimental approaches

Experimental and quasi-experimental evaluation methods are used to identify and measure the impact of an intervention. The main idea behind these methods is that their impact can be measured by comparing the 'counterfactual' outcome that can be observed from a 'control' group with that of the intervention (or treatment) group. For these methods to be implemented soundly, the treatment and control groups need to be identical (with the only difference being the intervention) or differ in ways that can be accounted for.

Analysing data from the treatment and control groups makes it possible to ensure that the changes that happened are a result of the intervention. It allows for the average additional or net change that is caused by an intervention to be quantified. The feasibility of an evaluation using an experimental or quasi-experimental method and the choice of the most suitable approach depends on factors such as whether an intervention can be randomised, the expected size of the effect and collected sample size, and the ability to use a control group (Magenta book, 2020<sup>47</sup>). Experimental impact evaluations with a counterfactual are mainly referring to Randomised Control Trials. Quasi-experimental impact evaluations (or QED) methods include matching methods, difference in differences, instrumental variables, timing of events, and regression discontinuity design; and these are used when randomisation cannot be undertaken.

### ■ Randomised control trials (RCTs)

When conducting a Randomised Control Trial, the evaluator can randomly assign individuals to the treatment group and the control group, an approach which allows them to estimate the causal impact of the treatment on participants. Given that participation does not follow self-selection into the project, this evaluation is net of the effects of both observable and unobservable characteristics of the individuals. However, there are many risks that might not allow for the evaluation to achieve its goals. High fidelity with the protocol (ie close/strict adherence to the planned model of implementation) should be maintained throughout the trial to ensure that the findings are legitimate.

<sup>45</sup> Ibid

<sup>46</sup> Ibid

<sup>47</sup> <https://www.gov.uk/government/publications/the-magenta-book>

### ■ Propensity score matching (PSM)

Matching methods are a form of quasi-experimental impact evaluation where a comparison group is collated from a secondary source of data. This method helps match the individuals in the sample to a group of individuals who were very similar to them prior treatment, but who have not participated in the programme. The two groups should be matched using parameters that affect both the probability of an individual participating in the programme and the intended outcomes of the programme. One caveat of this approach is that it can only match individuals based on their observable characteristics.

### ■ Difference-in-differences (D-i-D)

This approach allows for estimation of the effects of an intervention using a treatment and control group that are assumed to have similar trends over time in terms of the outcomes of interest. The control group is collected from a secondary data source. If the common trend assumption holds, this estimation method helps identify the causal impact of a programme on the treated group.

## Simple, complicated, and complex environments

According to the Magenta Book Supplementary Guide<sup>48</sup>, programme evaluation and policies can be distinguished into simple, complicated, and complex problems. An example of a simple problem is following a recipe, an example of a complicated problem is sending a rocket to outer space and that of a complex problem is raising a child. Policy interventions are almost never simple as they contain levels of complexity that increase according to the elements under consideration (different actions and organisations, for example). However, interventions can have different aspects some of which are simple, some of which are complicated, and some of which are complex. It is, therefore, more helpful to identify complicated/complex elements rather than classify an entire programme or intervention as complex. The work undertaken to develop a theory of change for the FLF programme (see Chapter X) indicates there is a degree of complexity, some in the intervention itself with its flexible and adaptable support, but mostly in the desired outcome and impacts. Therefore understanding the issues involved in choosing an evaluation method in complex is important.

Complex systems are difficult to define, but they share attributes such as many diverse and interacting components, non-linear interactions between components, and, in case of complex adaptive systems, adaptation or learning by the components in response to change. Complex systems also make the evaluation process more challenging (as indicated in the Magenta book) in various ways:

- They can be sensitive to context, making it difficult to define scope and boundary of the evaluation.

---

<sup>48</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/879437/Magenta\\_Book\\_supplementary\\_guide\\_Handling\\_Complexity\\_in\\_policy\\_evaluation.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879437/Magenta_Book_supplementary_guide_Handling_Complexity_in_policy_evaluation.pdf)



- They can make it harder to establish causal links (whether a certain intervention led to a certain outcome) due to the difficulty in unifying intervention and separating control groups.
- They may also result in different components having a disproportionate effect on the outcome, either by slowing down or accelerating change and it can also result in changes post-evaluation that are difficult to predict.

This does not necessarily mean that complex systems cannot be evaluated using experimental approaches, but they may be more suited for theory-based evaluations as they are better able to account for context.

There is no one size fits all for dealing with complexity. There are many evaluation methods available that account for and work well with complexity. The approach chosen will depend on the characteristics of complexity in the system, the purpose of evaluation (as discussed above) and the feasibility of the approach. Some guidelines for evaluation in complex settings include:

- system mapping that brings different groups together to share their vision,
- developmental approach that involves stakeholders especially in the adaptive management approach,
- qualitative theory-based approaches that can help capture mechanisms of change and offer lessons and information for similar contexts/policies,
- case-based approaches that help identify the combination of factors that are necessary for success in a certain context,
- computational system modelling that can provide a 'virtual' counterfactual when it is not possible to establish an experimental counterfactual.

Importantly, the details of methodological requirements may only become clear over time, meaning that evaluation design should be regularly reviewed to ensure that it is working well and to implement any modifications.

## Other dimensions

When thinking about different evaluation methods there are other dimensions to consider. Some methods, for example, are better used for learning purposes (answering why something works) and others are better suited for accountability purposes (what works in the first place). This can affect the choice of methods; if the purpose for the evaluation is accountability, then considerations such as external and internal validity may not be as important<sup>49</sup>. Conversely, if the purpose of evaluation is learning, then the method must be able to open the 'black box' of causality, meaning it is not enough to conclude that X causes Y, but to explain how X causes Y, through what mechanisms and under which circumstances. As with all dichotomies, learning and accountability are not necessarily

<sup>49</sup><https://repository.fteval.at/126/1/Broadening%20the%20range%20of%20designs%20and%20methods%20for%20impact%20evaluations.pdf>



mutually exclusive; some evaluation questions and methodologies can encompass both these purposes<sup>50</sup>.

Another dimension is the type of data that the methods require to be collected and analysed; for some methods, it is more appropriate to rely on quantitative data, while in others this may not be possible or desirable, so interviews, focus groups, and other qualitative methods are used. In most cases, the evaluation process will require the use of mixed methods approaches, whereby qualitative and quantitative data are collected to serve different purposes or to shed light on different aspects.

## Choosing an evaluation method

The expansion in evaluation methods in the past few years has made the challenge for designing an evaluation less about the scarcity of methods and more about choosing from an abundance of methods. Shortlisting methods has become a challenge as each approach has its own requirements for feasibility, comparative advantages, and shortcomings.<sup>51</sup> When selecting an appropriate approach, one needs to consider the practical elements of the choice such as time and budget constraints and availability of data, as well as the design elements to the choice such as does it answer the evaluation question and is it appropriate for the characteristics of the intervention. Different methods will require different resources, timescales, and may or may not allow for iterations, some methods will be better at answering questions, some will be more-cost effective, and some may be quicker<sup>52</sup>.

The evaluation literature highlights that selection of the analytical method should be informed by the Theory of Change and the uncertainties and assumptions that it identifies; and that stakeholders are involved in the selection process as well as the development of the Theory of Change to ensure that the outcomes they are interested in are captured, that they understand the limitations and what is feasible. This is to ensure that implications, no matter the direction, are accepted. It is important to also clearly define the population, the place and the time period of the evaluation, as well as the purpose of the findings, and whether or not they can be generalised<sup>53</sup>.

Most evaluations will use a mix of methods, both qualitative and quantitative, for different purposes. The process is also likely to be iterative, with learning and reviewing occurring during the evaluation. No method will be perfect, but some will have advantages over others in terms of appropriateness, timing, and cost<sup>54</sup>.

---

<sup>50</sup> *ibid*

<sup>51</sup> [https://www.bond.org.uk/sites/default/files/caem\\_narrative\\_final\\_14oct16.pdf](https://www.bond.org.uk/sites/default/files/caem_narrative_final_14oct16.pdf)

<sup>52</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/879438/H\\_MT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/H_MT_Magenta_Book.pdf)

<sup>53</sup> *ibid*

<sup>54</sup> *Ibid* & [https://www.bond.org.uk/sites/default/files/caem\\_narrative\\_final\\_14oct16.pdf](https://www.bond.org.uk/sites/default/files/caem_narrative_final_14oct16.pdf)

The Design Triangle was developed by Department for International Development<sup>55</sup> as part of a broader report on designs and methods of impact evaluations. It frames the method choice as a function of interacting aspects of the evaluation. Especially in the context of complex environments, it is important to analyse the programme's or the intervention's attributes and to understand what that means for the evaluation designs and methods. Choosing an appropriate evaluation method will require aligning evaluation questions – which is the core of any evaluation – with available designs and the specific characteristics of the programme.

In essence, it is important to ensure that the evaluation questions and characteristics are framed correctly and understood well before choosing a method. The choice will depend on what is appropriate in a particular situation, considering the nature of what is being evaluated, the limitations, who the evaluation is for, and for what purpose. It is important to account for as much as possible from the beginning and adapt to changes and details as they come<sup>56</sup>.

---

55

<https://repository.fteval.at/126/1/Broadening%20the%20range%20of%20designs%20and%20methods%20for%20impact%20evaluations.pdf>

56 <https://www.betterevaluation.org/en/choose-methods-and-processes>

## Appendix D: Learning from other evaluations

---

The small number of programmes that have been evaluated and that have similar aims to the FLF programme are noted below. The evaluation reports were reviewed to extract details of the evaluation method adopted including (where appropriate) measures and data sources used and challenges encountered, and to identify any lessons for an evaluation of the FLF programme.

### UK programmes

1. **Royal Society of Edinburgh Enterprise Fellowships** (BiGGAR Economics, 2019). The Royal Society of Edinburgh Enterprise Fellowships supported 244 fellows between 1997 and 2017. It is a programme targeted at entrepreneurial researchers and innovators who want to begin their own business. The support is financial and non-financial; £55,700 are spent on each fellow on average, while fellows benefit from interactions and support provided by the Royal Society of Edinburgh (RSE), university partners, UKRI and others. The programme lasts one-year and the support package consists of salary cover, business training, business support funding, travel expenses for training, mentoring and advice and membership of the RSE Entrepreneurs' Club.

The evaluation<sup>57</sup> of this scheme consisted of a survey directed only to the scheme participants and case studies. This qualitative evaluation looked into how the fellows were supported, limitations and possible enhancements. Without a comparison group and measurable outcomes, this evaluation helped understand how the programme supported the participants but did not give a clear understanding of its impact.

2. **Clinical Research Training Fellowship** (Stewart, et al., 2012). The Clinical Research Training Fellowship provides support for up to three years to clinically qualified candidates to undertake specialised or further research training in biomedical sciences so that they can acquire a higher degree (for example a PhD or an MD). This programme aims to support the development and progression of Clinical Academics by supporting them so that they can undertake postgraduate research training.

In this evaluation<sup>58</sup> alumni fellows were contacted and were asked to answer a survey. Two cohorts were targeted; the first one was awarded the fellowship in 1991 (40 respondents) and the second was awarded the fellowship between 1993 and 2003 (191 respondents). This study did not have a control group and produced descriptive statistics based on the responses of the survey respondents. Even though the survey and the use of two different cohorts rather than one offered some insight into the effects of the

---

<sup>57</sup> <https://www.rse.org.uk/rse-publish-evaluation-and-economic-impact-assessment-of-enterprise-fellowships/>

<sup>58</sup> <https://bmjopen.bmj.com/content/2/4/e001792>

scheme on the candidates, the lack of a comparison group means that it is not possible to draw clear conclusions about the programme's impact on its Fellows' careers.

3. **CLAHRC East of England Fellowship Evaluation** (The National Centre for Post-Qualifying Social Work and Professional practice, 2018). The CLAHRC is a programme designed for clinicians, health and social care practitioners and managers to undertake and disseminate high quality research, to build capacity, facilitate collaboration between HEIs and NHS organisations, to bridge the translation gap between research and practice, and bring about sustainable improvements to the delivery of services. There were originally 9 locally-based programmes, increasing to 13 in 2013. NIHR in the East of England commissioned Bournemouth University to evaluate their Fellowship. The ambition for the fellowship is that it is transformative – increasing the skills of practitioners but enabling them to act as leaders and agents for change and further supporting integrated knowledge transfer. Fellowships last for one year and the model provides funding to allow for one day a week to be dedicated to the programme; and monthly teaching workshops, bi-monthly action-learning sets, and on-going academic support through supervision.

The evaluation<sup>59</sup> sought to measure long term behavioural change and 'identifiable outcomes.' It identifies three levels of impact: primary – individual impact, secondary – service, team, organisation or local policy, and tertiary – service users, national policy. The evaluation took a theory-driven approach and gathered data from fellows using questionnaires and telephone interviews (N=29 and 12 respectively) and third party testimonies (as verification, N=12). It appears that there is no longitudinal element, no comparison group and no secondary data sources used. Measures (captured in the survey) included reflections of impact on their wider professional practice with patients, colleagues and policy; engagement in further research (after the ending of the Fellowship) noting published outputs, service changes/improvements, professional qualifications; and also future intentions to participate in research.

4. **ESRC Postdoctoral Fellowships** (Meagher, 2004). The ESRC Postdoctoral Fellowships are directed to recent PhD graduates in Social Sciences, which means that people can only apply within one year after graduation. It supports fellows by offering them guidance (by appointed mentors) for them to develop their research agenda and research skills, publish their work, disseminate their work to wider audiences, and finally, secure university appointments. The programme has been running since 2001 and an evaluation took place in 2004. By that time 293 fellowship awards were granted. The targeted audience also included some priority areas within Social Sciences. Those were areas with relatively low numbers of PhD graduates who secure academic posts following their doctoral studies.

The evaluation<sup>60</sup> focused on capacity, publications, skills, and development of fellows. In terms of capacity, the main four outcomes were whether the fellows secured academic posts, whether the fellowship affected their job prospects in general, how those

<sup>59</sup> <https://www.clahrc-eoe.nihr.ac.uk/wp-content/uploads/2018/05/reducedx2-ecopy-CLAHRC-Fellowship-Evaluation.pdf>

<sup>60</sup> <https://esrc.ukri.org/files/research/research-and-impact-evaluation/esrc-postdoctoral-fellowship-scheme/>

outcomes varied between priority fellows and other fellows, and finally whether good non-UK researchers were kept in the country. In terms of publications, both quality and quantity were included as outcomes of interest. The skills and development outcomes focused mainly on the development of new methods through advanced training and on engagement with non-academics. Finally, the evaluation made recommendations on how the scheme could become better.

The report followed a mixed-methods approach where over 80 report of past fellows were studied and analysed; 45 discussions with former fellows, mentors, and Heads of Departments took place; questionnaires to mentors were distributed; and one focus group with fellows was conducted. The results drawn from those data sources were informative, however they relied heavily on perceptions of the individuals involved and no counterfactual group was used to set a baseline. Furthermore, the comparisons focusing on the effects of the fellowship on priority groups versus other fellows might not have represented the actual effects of the fellowship on those priority groups as fellows from priority groups might be heavily selected from the sub-group that would have followed an academic career regardless.

5. **EPSRC Postdoctoral Fellowships** (unpublished). The EPSRC Postdoctoral Fellowship is aimed at individuals who hold a PhD or have equivalent research experience and allows them to conduct novel and world-leading research. The fellowship covers 80% of the project costs while the rest is contributed by the hosting institution. The grant can be used for staff costs, equipment, costs related to impact and travel.

There is an ongoing evaluation of EPSRC Fellowships including early career grants and also grants directed towards researchers at a more advanced stage of their career. The evaluation focuses on fellowships awarded since 2006, they have duration spans up to five years, and the fellows spend between 50% and 100% of their time on the fellowship. The outcomes of interest with respect to the effects of the fellowships on the fellows' careers are publications, dissemination, awards, other funding, collaborations, mobility between institutions, impact and spin-outs. The method includes the collection of information on 500 to 600 EPSRC fellows from Scopus and ResearchFish. A survey is also rolled-out to alumni fellows to collect their views on their fellowship. Finally, semi-structured interviews with alumni allow the exploration of the personal career impacts of those grants. This is a very comprehensive collection of information that helps understand the fellows' career paths and their experience with EPSRC, but it seems that there are no plans to include a counterfactual group in the analysis.

6. **Independent Research Fellowships (NERC)** (Gleed & Bennett, 2018). The NERC Independent Research Fellowships are early career grants for PhD holders in environmental sciences and help participants build research independence. These grants are directed to early career postdoctoral researchers within eight years from their PhD certificate date. The duration of the fellowship is five years, full-time.

This evaluation<sup>61</sup> took place in 2017 and its main aim was to understand whether the training and support opportunities available to grant holders were sufficient in achieving

---

<sup>61</sup> <https://nerc.ukri.org/about/whatwedo/engage/engagement/ecr2017/ecr-evaluation/>

the goals of the fellowship. The methodological approach consisted of a survey which was directed to 329 ECR grant holders and 69 EC employers. This was followed up by ten interviews with survey respondents, which translated into five case studies. This approach offered a good understanding of the characteristics of grant holders and their experience in the programme and led to a series of recommendations. There was no counterfactual group involved so the results were not identifying the effects of the support given on measurable outcomes.

7. **David Phillips Fellows and Discovery Fellowships (BBSRC) (UKRI, 2020).** The BBSRC fellowships support researchers in biosciences to become future leaders. The Discovery Fellowship (formerly Anniversary Future Leaders Fellowship) is for early career researchers and supports them while they conduct independent research in a host lab. The David Phillips Fellowship is directed to more senior researchers and supports them with establishing their own independent lab. The discovery Fellowship was introduced in 2014 and the David Philips fellowship in 1992.

The evaluation<sup>62</sup> of those two BBSRC fellowships took place between 2019-2020. It focused on how those fellowships led to the fellows' personal development, with respect to their research outputs, promotions, permanent contracts, and more senior fellowships. The methodological approach involved 13 interviews with Discovery Fellows and 7 David Phillips Fellows; analysis of ResearchFish data at various stages of progression through their award; starter and finisher surveys with 58 starter and 20 finisher Discovery Fellows and 19 starter and 6 finisher David Phillips Fellows; and finally analysis of the applications received (successful and unsuccessful) for those fellowships and success rate by characteristics such as gender, ethnicity, and career stage. This approach did not have a counterfactual group and a causal estimation of the effects of the fellowships on outcomes.

8. **Dorothy Hodgkin Fellowship (DHF) and University Research Fellowship (URF),** The Royal Society (Mellors-Bourne, et al., 2018). Those are two long standing fellowship programmes aimed at early career researchers, running for 26 and 38 years, respectively. The URF provides up to ten years of funding (80% salary costs, contribution to research expenses), support (training, cohort activities, facilitated public engagement, industry engagement and outreach) and dedicated time for research for ECRs with the potential to become leaders to build an independent research career, and has supported approximately 1600 individuals (up to 2018). The DHF offers four to five years of funding, support and dedicated time for those earlier in their career and needing more flexibility with recipients tending to be women with young families and has supported over 200 individuals. Together 50 to 60 new fellowships are offered each year.

CRAC (with support from IES) were commissioned to determine the impact of the fellowships on career pathways, the scientific achievements and contributions of the alumni including leadership roles, influencing policy, commercialisation, and public engagement<sup>63</sup>. The study also sought to gather evidence to improve the fellowship

<sup>62</sup> <https://bbsrc.ukri.org/documents/review-of-fellowships-investments/>

<sup>63</sup> <https://royalsociety.org/grants-schemes-awards/career-pathway-tracker/>



programme and influence research culture. The work involved a survey of current and past fellows. It is not a true impact study as no control/comparison groups were established, but where possible aggregate results were compared to the Principal Investigators & Research Leaders Survey (PIRLS). The survey explored current experiences: location of employment (academia or elsewhere), geography of employment, contract status, international collaborations; and career progression over time after completion of the fellowship in terms of job roles and time taken to achieve these, supervision, management and training responsibilities (including numbers of students/staff), time spent abroad, regularity/extent of public engagement, commercialisation activity, knowledge exchange activity, contribution to national/international policy. Virtually all fellows pursued a career in academia, there was some movement over careers in and out of academia, but only 5% of URF and 12% of DHF worked outside of academia in commercial, third sector or policy/funding bodies attaining senior leadership roles and engaging in public engagement, commercialisation activity, and policymaking. Indicators used as proxies for establishing independent research leadership included: publish a key paper as a PI, secure significant research grant as PI, supervise at least one doctoral student to successful completion, hired at least one postdoctoral researcher, obtained permanent academic position. These reflect the dominance of academic career pathways among fellows. Challenges encountered in undertaking the analysis included: changes to eligibility requirements and definitions and change in duration of fellowships.

9. **Clinician Scientist Fellowships**, Health Foundation and Academy of Medical Sciences (Jenkins & Bryant, 2012). Launched in 2001, this competitive scheme is aimed at outstanding professionals, mid-career researchers in academic or clinical settings to pursue and develop research in a clinical setting. Its ambitions are that fellows will cultivate important research programmes, create and lead research teams, and become the next generation of clinical academic leaders. To date it has supported 26 fellows, selected over four rounds. The scheme offers five years funding (averaging at £550k per fellow for the scheme) and support including a leadership development programme, mentoring and networking opportunities. Eligibility requirements include medically qualified (PhD or MD in basic science or clinical/health-related subject), approaching or recently awarded consultant status.

An independent evaluation<sup>64</sup> was undertaken by Jenesys Associates in 2011/12 (after 21 awards had been made across three cohorts) and an evaluation report was published by AMS in 2013. It follows on from an earlier evaluation in 2005 (of 13 fellows). The evaluation approach involved: an online self-report survey (unvalidated) of fellows using indicators from previous evaluations, and interviews with fellows and other high-level stakeholders to explore impacts and benefits/strengths of the scheme and potential improvements. There was no comparison or counterfactual group. The evaluation was structured into: research impacts (measured through number of publications in significant journals, prestigious prizes, leveraging additional funding to create/grow research groups and develop projects, number of new research collaborations (including

<sup>64</sup> <https://acmedsci.ac.uk/file-download/35282-CSFSchem.pdf>



inter-disciplinary and cross sectoral collaborations), teaching and supervisory activities); clinical and healthcare impacts (new protocols and guidelines, clinical trials of new products, development of new techniques or devices, successful filing of patents in key fields, and increased research awareness and application in NHS, encouraging evidence-based practice); leadership and career impacts (retention within clinical academia, promotion to senior fellowships and tenured appointments, collaborative research funding applications, establishing own teams/influencing work of others, plus encouraging research ambitions, enhancing confidence, facilitating access to key networks, and becoming perceived as emerging leaders). The evaluation also looked at how the scheme delivered impact, what were the key factors about the scheme that led to its success (impact).

Overall the study found: 'CSF fellows are being enabled to advance knowledge and establish themselves as a leader in their field, with 297 significant publications and 37 prestigious prizes reported across all cohorts. It is also building research capacity and CSF fellows have leveraged over £50 million in additional research funding across all cohorts, which is a return of £4.45 for each £1 invested in the scheme. CSF fellows are also forming important collaborations, at the inter-disciplinary level, and between industry and the NHS. The evaluation has also found evidence that the quality and performance of healthcare is improving as a result of the fellows' research, with research being translated into clinical practice and healthcare policy at local, national and international scales. This has been manifested through the optimisation of clinical protocols and guidelines, the successful filing of patents, and the development of new clinical trials, devices and technologies.' (foreword)

A number of challenges were noted relating to the differential impacts identified: influence of the type of research on lead time to impact, time since award was made/length of career.

10. **Springboard**, Wellcome Trust and others (Freshney Consulting and Aleron, 2019). The programme was launched in 2015 and to date has made 105 awards (from 396 applications, a success rate of 27%). The scheme offers funding (up to £100k which can be used flexibly) and support over two years, including access to mentoring and career development. It is aimed at non-clinical biomedical researchers in academia at the start of their first independent post (eg lecturer, junior group leader) to help them launch their research careers, and is flexible allowing individuals to take career breaks. It is open to those in their first independent/salaried position (up to four years after appointment), but they must not already hold grant funding for more than £150k. A similar award was identified as Wellcome Trust's Seed Awards (although has a broader target)

An evaluation<sup>65</sup> by Freshney Consulting (specialists in supporting medical research) and Aleron was undertaken for the Academy of Medical Sciences and published in 2019. This sought to evaluate the contribution of the scheme in advancing careers of Biomedical Scientists, and only one of five objectives related to understanding the impact of the scheme on individuals and their careers. The aspects explored included:

<sup>65</sup> <https://acmedsci.ac.uk/file-download/85493>

research outputs, citation analysis, funding applications, perceived value and impact, career progression, mentoring, and collaborations /networking and engagement. Research output analysis used ResearchFish (data provided by awardees) and an applicant survey. The Citation Analysis used data from Web of Science, provided by Clarivate Analytics. Insight on scheme value and experience used an Online survey of successful applicants and unsuccessful applicants. The methods used included some degree of comparison group. The evaluation included: a survey of awardees and a survey of unsuccessful applicants; surveys of wider stakeholders (scheme champions, panel members); a small number of interviews/case studies; and citation analysis with comparison of awardees against unsuccessful applicants. To draw a comparison group of unsuccessful applicants, a random sample of 19 and 20 (equal to the number of awardees in each round) was generated. The citation analysis used ORCID numbers of both groups to extract relevant data from ResearchFish supplemented by PubMed, ORCID website, and awardees own websites/research pages, and publication window was set for each cohort. Clarivate Analytics was contracted to carry out the citation analysis using the Web of Science platform to identify the number of citations for each paper and the number of citations for each individual (awardee and unsuccessful applicant) from 2016 to 2019. The authors heavily caveat this analysis as intermediate results as it can take 3 to 5 years for citations to accrue on published papers (and only four years was analysed) and note that the authorship position was not taken into account.

The evaluation assessed: research outputs (awards/recognition factors, new collaborations and partnerships (academic, clinical or industrial), new research grants, new research publications, public engagement activities, funding secured (expressed as a total, and as a return for each £1 of programme funding), grant applications submitted, size of grants applied for, research papers published, citations received); value of funding and funding leverage (additional funding from Institution (eg to appoint PhD student), promotion, dedicated time for research, teaching time, personal recognition and confidence).

The evaluation also included some mapping of major funding programmes for non-clinical biomedical researchers and mapped programmes against the MRC's careers framework which has five career stages: consolidation, exploration, progression, independence, and leadership. It identified four groups of programmes: postdoctoral awards (NIHR Advanced Fellowship, Sir Henry Wellcome Fellowships, BBSRC David Phillips Fellowship, Medical Research Charity Fellowships); Transition to independence awards (UKRI FLF, MRC Career Development Awards, Sir Henry Dale Fellowships, Dorothy Hodgkin Fellowships, MRC New Investigator Research Grants, Springboard Awards, Wellcome Trust Seed Awards, Medical Research Charity Fellowships); Skills Development Awards (MRC Skills Development Fellowships, NIHR Development Fellowships, NIHR Development and Skills Enhancement Award); Research/project grants (MRC Research Grants, BBSRC New Investigators, Wellcome Trust Investigator Awards).

11. **Engineering for Development Research Fellowship**, Royal Academy of Engineering (CRAC, 2020). RAEng run a number of research programmes designed to promote

excellence in UK engineering research and innovation and enhancing partnerships with industry. They are aimed at different career stages, with Research Fellowships targeted towards early career researchers with the objective to support them in establishing their independence and international reputation, to pursue an ambitious programme of engineering research, and to become ambassadors for the Academy and advocates for STEM disciplines. Fellows receive up to £500,000 over a five-year period, plus additional support in the form of mentoring support, training opportunities, networking with other Fellows, and dedicated time for research. Research Fellowships are open to individuals of any age and nationality, but they must be based in a UK HE institutions, propose an engineering-focused project, and must have the support of their host institution (who will employ them and permit them to devote all their working time to the Fellowship programme. Applicants must have a PhD, which was awarded no more than four years prior to application to the Fellowship and must not hold a permanent academic position before the start of the Fellowship. The Research Fellowships were launched in 2001, with an application round each year. This is currently running at about 100 applications per year for 16 to 18 awards.

CRAC (with support from IES) were commissioned in 2019 to undertake an evaluation and career tracking of 4 programmes including the ECR Research Fellowships<sup>66</sup>. This sought to determine the long-term impacts of the programmes on careers and contributions to engineering research and industry and understand more about career pathways and progression of participants and their wider contributions. The work also included a process evaluation element to consider programme operation and potential improvements, and a future focus to devise an approach to facilitate ongoing career tracking of alumni.

The approach involved: consultation with stakeholders, review/analysis of programme monitoring data, an online survey of current awardees and alumni capture experiences, career progression and perceptions of impact, and follow-up indepth interviews with a sample of alumni and industrial partners. The survey was sent to 110 current award-holders (starting prior to 2018) and alumni, and 66 responses were achieved (54 alumni and 12 current award-holders). The survey captured:

- subsequent employment (after completing the award),
- research related outputs (including: published peer-reviewed papers as lead or contributing authors, established international collaborative research, helping to train doctoral researchers, supervising postdoctoral researchers, keynote conference presentations, significant public engagement activity, established collaborations with industrial partners, undertaken outreach work, developed patents or other IP assets, developed new products including software, senior-level policy making, setting up a company),
- reflection of fellowship's impact on their career to date (including: faster career progression, easier to secure permanent position, level of seniority reached, enhanced perception by colleagues, establishing independence, establishing

<sup>66</sup> <https://www.raeng.org.uk/publications/other/raeng-research-programmes-evaluation-final>

international reputation, pursue new directions, stronger publication record, success in further grant funding, improved self confidence, establishing new collaborations, gaining expert knowledge, gaining new skills/technical competencies, enabling multi-/cross-disciplinary work, opportunity to work with industry),

- perceptions of 'additionality' of fellowship to career impact

It is not a true impact study as no control/comparison groups were established. Also, the research was undertaken with cohorts from across several years including those who completed the fellowship some time ago (with no baseline established at the start of the programme).

12. **Rising Stars**, Sêr Cymru (the Welsh Government Office for Science) (Bryer, et al. 2018). This programme was aimed at mid-career researchers with over 7 years' experience from completion of their PhD (or equivalent) and with a scientific track record showing great promise. Applicants could be in academia or business but required the support of their host organisation. The support package included: £200k per year, and 10 awards were made (against a target of 26). This programme (along with 30 Research Fellowships aimed at those 3-5 years post PhD and providing 3 years' support, and 12 Recapturing Talent Fellowships aimed at those who had had a career break) has been replaced with Future Generations Industrial Fellowships. These Future Generations Industrial Fellowships are offered alongside a number of Accelerator awards (Strategic Partnership, Infrastructure, and Capacity Building Accelerator Awards), and form part of the Welsh Government's Office for Science' approach to increasing research capacity in STEMM related subjects (although the programme is open to researchers working in relevant areas of applied social science). Collectively this is the Sêr Cymru II initiative, which was launched in 2012.

The Rising Stars and Recapturing Talent programmes were evaluated<sup>67</sup> by OB3 and Regeneris as part of the Sêr Cymru II programme mid-term evaluation published in 2018<sup>68</sup>. This found overall that 51 fellowship awards were made (against a target of 56) across four rounds: 33 for Research Fellowships, 3 for Recapturing Talent, 9 for Rising Stars, and 6 for Chair funding packages. The evaluation looked at the progress made by fellows on several aspects, but concluded it was too early to see any meaningful change.

The approach taken to the mid-term evaluation included: desk-review of performance and impact (review of programme monitoring data and progress against targets, stakeholder interviews, analysis of secondary data using HESA data), a web-based

<sup>67</sup> <https://gov.wales/sites/default/files/statistics-and-research/2019-01/evaluation-of-the-ser-cymru-ii-programme-mid-term-evaluation.pdf>

<sup>68</sup> An initial, inception evaluation in 2017 had involved: key stakeholder interviews, literature review, development of a Theory of Change model, review of programme monitoring arrangements and establishing a baseline position.

[https://www.academia.edu/35078571/S%C3%AAr\\_Cymru\\_II\\_Inception\\_Evaluation](https://www.academia.edu/35078571/S%C3%AAr_Cymru_II_Inception_Evaluation)

survey of all funded research fellows, and 11 case studies of fellows (involving interviews with a sample of fellows and their academic supervisors and collaborating industry representatives).

The progress of fellows was measured using the survey, supplemented with interviews, and this explored: fellowship promotion and application process, fellowship implementation and reporting requirements, induction, training and support provided, collaborations with industry and academia, and early outcomes and anticipated impacts of their research. The latter included levels of grant funding applied for and secured, number and range of public engagement activities, submission of papers, and publication of papers. The survey received 15 responses.

The authors note how it was not possible to undertake fieldwork with unsuccessful applicants during the mid-term evaluation as no data sharing consents had been secured to share applicant data with the research team. Thus, critically, the interim evaluation recommended that a final impact evaluation be conducted during the last six months of the programme, and a brief counterfactual impact evaluation update be commissioned 12 to 18 months after the programme ends (to allow for published data to be considered). Also, that appropriate data sharing agreements are put in place to allow this impact evaluation to consider feedback from successful and unsuccessful applicants, plus collaborating businesses and third sector organisations.

It is not yet a true impact study as no control/comparison groups were established. Also, the numbers involved are very small and the response rates were small.

### Innovator, leadership and knowledge transfer schemes

13. **ICURe** (Ipsos MORI, George Barrett & Tomas Ulrichsen, 2018). The ICURe scheme aims to boost entrepreneurial skills and commercial awareness within academia by supporting commercialisation of academic research to be taken forward by academia and industry. Funding and training is provided to university researchers from all academic fields with potential commercially relevant research outputs. This programme was introduced in 2014.

The evaluation<sup>69</sup> focused on the six first cohorts between 2014-2016 and looked at the effects of the programme on entrepreneurial skills and commercial awareness as well as other wider market relevant effects. The evaluation used a mixed methods approach including the analysis of application data (of all applicants) and monitoring data (of the successful applicants); stakeholder interviews; 10 case studies of individual projects; a survey; econometric analysis of individual level data; and descriptive analysis of aggregate level data. The analysis of application and monitoring data helped understand who applied and the outcomes achieved by the successful applicants; the stakeholder interviews showcased the issues that institutions face and how the programme works within that context; the case studies helped understand the particulars of the programme

69

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/673319/ICure\\_Evaluation\\_Report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673319/ICure_Evaluation_Report.pdf)

implementation; the applicant survey had a sample of 163 programme participants and 59 programme applicants who were not awarded a place in the programme. The survey was used in the quantitative analysis investigating the effects of the programme on the outcomes. Difference in differences was used to analyse the effects of the programme. Finally, descriptive analysis of secondary data from the Higher Education-Business and Community Interaction Survey helped understand the wider context on commercialisation activity in institutions participating in the programme.

This evaluation had a very comprehensive approach where the wider context was described and the impacts of the programme on participants were estimated. The difference-in-differences helped identify the effects of the programme to the extent that the observed and unobserved characteristics of participants and unsuccessful applicants were properly controlled for. It was noted that the non-successful applicants were at an earlier stage of their career compared to the successful applicants which means that the estimated positive effects of the programme might have been overstated. Furthermore, if successful applicants were also more likely to participate in other similar programmes during that same period without this being properly controlled for, then that would have overstated the effects of the programme even further.

**14. DigitalHealth.London Accelerator** (Heath Innovation Network) (KADA Research 2020). This is a £3.4m programme delivered by London's three Academic Health Science Networks, MedCity and Chelsea and Westminster Hospital NHS Foundation Trust. It provides tailored support from NHS professionals for SMEs in the digital health sector aiming to enter/navigate the NHS market, to help them develop, test and pilot their innovations within the NHS. Four cohorts (2016 to 2020) of SMEs will be supported for a period 12-months each (a longer period than offered by many other accelerators). The support included access to a named relationship manager, networking and educational events, bespoke diagnostic and brokerage meetings, and access to mentor networks enabling knowledge building (eg about NHS decision-making) and making connections. The programme aims to enhance London's competitiveness by tackling market failures related to poor collaboration and coordination, complexity of innovating in the NHS, and challenges translating knowledge into new market products and services.

The programme has been subject to an interim independent summative assessment, by KADA research, of its performance, benefits and impacts including impacts on the London health economy. The evaluation<sup>70</sup> also aimed to assess efficiency, effectiveness and value for money, factors contributing to success and failure, whether the programme met businesses expectations/added value, and areas for programme improvement. At the time of the interim evaluation- 75% of the budget had been spent, three cohorts (out of four) completed, 97 businesses supported (total expected number will be 113 over the target of 105) and 553 applications made, a success rate of 18%

The evaluation used qualitative and quantitative approaches drawing on programme reported outputs and spend, and a survey with SMES – 21 beneficiaries and 1 counterfactual. The approach to impact assessment appeared to be: a) survey of

<sup>70</sup> [https://digitalhealth.london/wp-content/uploads/2020/09/Digital-Health-Accelerator\\_Interim-Report\\_v1-2.pdf](https://digitalhealth.london/wp-content/uploads/2020/09/Digital-Health-Accelerator_Interim-Report_v1-2.pdf)



recipients (21 beneficiaries and 1 counterfactual, all participants SMEs were approached but just those willing to be surveyed were contacted so it is a self-selecting sample), b) analysis of programme monitoring data (including reported outputs and collection of economic impact metrics via a survey undertaken by the delivery agent), and c) case studies. There was no secondary data used and no counterfactual assessment (beyond asking recipients if commercial benefits would have happened without the programme). The business survey however suggests some areas that could be investigated in a survey of FLF hosts.

- The business (telephone) survey (completed by 21 supported companies) looked at motivation, satisfaction, and impacts. It covered: professionalism of Accelerator staff, whether expectations for the programme were met, perceived reduction in barriers to innovation, perceived 'closeness' to the market (before and after programme), and improvements made as a result of the support (additionality in progress made). The latter included job creation, R&D spend, products or services developed, and additional company turnover resulting from the programme measured in bands, all measured at the time of the survey and future expectation; any other outcomes in terms of new sales/customers, internal expansion (recruitment onto project/internal investment), further private investment, and further public investment; other innovation and commercial benefits including opportunities to showcase new products/services, promote business investment in R&I, new pilot projects, new contacts, enhanced cooperation with research entities, support research/early product validations/advanced innovations; and any wider NHS/health benefits including opportunities to benefit patients, NHS savings and impact on high unmet medical needs and improvements in care commissioning. The survey asked respondents about additionality/deadweight – whether the commercial benefits made resulting from the programme would have occurred anyway with specific answer categories (not at all, at a later date, by a smaller amount, later and a smaller amount, same way). The survey also gathered feedback on the application process, and areas for improvement. An economic impact assessment was undertaken (using HM Treasury Green Book principles) to calculate value for money and net impact using the survey evidence – this took account of direct employment and indirect employment (eg on suppliers using a composite multiplier), deadweight, (using survey findings), displacement and leakage (using average figures)
- The programme monitoring data included: capital expenditure, SMEs receiving support, SMEs receiving non-financial support, SMEs collaborating with research entities, SMEs supported to introduce new market products, and SMEs supported to introduce new to firm products. In all cases the target and actual numbers were captured. The survey run by the Accelerator captured data on whether: companies had grown, additional employees attributed to the programme, additional contracts gained attributed to the programme, additional investment raised attributed to the programme, export contacts attributed to the programme. Plus whether the proportion of additional pilots attributed, additional NHS savings attributed, patients benefiting attributed.

The evaluation estimated the programme spend of £2.5m had resulted in £36.7m of GVA, generating £14.50 for every £1 spent. It estimated that 513 gross jobs resulted



from the programme. The evaluation also monitored equality – gender, ethnicity, and disability of business founders.

15. **Knowledge Transfer Partnership (KTP)** (Ternouth, et al., 2012). KTP is a key initiative in the UK with the explicit aim of fostering university-industry knowledge. The official description of the KTP is:

*“Support to UK businesses wanting to improve their competitiveness, productivity and performance by accessing the knowledge and expertise available within UK Universities and Colleges. KTP establishes a relationship between a business and an academic institution which facilitates the transfer of knowledge, technology and skills to which the business partner currently has no access. Each partnership employs one or more recently qualified people to work in a business on a project of strategic importance to the business, whilst also being supervised by the Knowledge Base Partner.”*

One of the largest formal evaluations of KTP was undertaken by Ternouth et al (2012). This group adopted a multi-faceted approach including the following:

- Identify key bodies of the literature
- Construct a generic KT model from this extensive review
- Undertake a quantitative analysis of completed KTP projects
- Undertake primary qualitative research with selected partnerships
- Synthesize the results of the research and compare the KTP model as it operates in practice with the generic model built inductively from the literature to draw conclusions regarding the KTP process and make recommendations

The research concluded that KTP success is contingent upon the integration of the structural and human capital processes inherent in the implementation of knowledge transfer processes. A specific aspect of this process was identified as being able to turn ‘potential absorptive capacity’ into ‘realised absorptive capacity’. Importantly, absorptive capacity is cumulative and individuals and businesses learn-by-doing thus creating a greater potential to undertake successful open innovation. It is this realised potential supported by enhanced absorptive capacity (the ability to absorb knowledge and use it) that creates superior economic outcomes (jobs, growth and productivity) that have been identified for innovation activities.

16. **High Potential School Leaders (HPSL) programme** (BMG, 2015). Established in 2006 by the National College for Teaching and Leadership (NCTL), the HPSL programme is a leadership development programme designed to raise levels of pupil achievement in challenging schools by developing high-potential school leaders to become head-teachers in these settings. Ultimately it aims to improve the life chances of pupils from disadvantaged backgrounds through outstanding school leadership and management. Initially, graduates of the HPSL programme were expected to take up a headship position in a challenging school within 4 years of commencing the programme, though this expectation is now expressed as a 2-5 year period.

A formal interim evaluation conducted by BMG in 2015 aimed to provide a picture of how programme participants were using the skills and knowledge gained through the programme to improve leadership and attainment within challenging schools; how past participants have done this; and what impact this has had on schools and pupils (including both perceived impact and where impact can be evidenced through available data, wherever possible the research distinguishes between these types of reported impacts).

In the scoping phase, a logic model which included key research questions was co-designed with NCTL. This logic model was intended to guide the formal evaluation process and focus. This formal element had 2 distinct phases which logically followed on from each other. However, each phase included the same two methodologies for consistency. These two methodologies used were: (1) a desk review of impact evidence; and (2) qualitative research with current/past participants and Residency Heads. In total across the two phases, 108 in-depth case studies were achieved, 5 on-line focus groups, and 100 impact initiative forms.

Phase 1 outcomes – On the delivery aspects and content of the residency year, the Residency Heads and HPSL participants interviewed in Phase 1 indicated a significant level of satisfaction for the support they were provided with and accessed through the programme. The study tour was seen to be particularly effective and beneficial element as it provided more practical exposure to different types of schools and approaches. Experiences with mentors and LDA coaches varied depending on the length of time available to spend with them and the quality of the relationships that were developed.

Phase 2 outcomes - Most of those interviewed strongly believed that the overall content and delivery of the programme was of a high quality; in particular participants valued the support received from their Learning and Development Advisor (LDA) and the networking opportunities the programme provided such as: contacting and exchanging ideas with fellow participants during and after they completed the programme and networking with professionals from other schools. One of the key ways in which the HPSL programme aimed to have an impact on schools was by requiring participants to undertake an 'impact initiative', a project designed and delivered by the participant in their school to achieve a positive outcome. Analysis by BMG of the Impact Initiative Forms which set explicit targets for establishing outcomes at the school level found that nearly two thirds (65%) of the 100 participants had met or exceeded at least one of their two specified targets; a further 32% had partially achieved at least one of their targets. In contrast, only 3% had not achieved or partially achieved any of their targets.

## Wider programmes

### 17. Marie Skłodowska-Curie Actions (many schemes) (European Commission, 2017).

This is a large-scale programme that runs across many countries. The most recent leg run between 2014-2016 and it focuses on promoting research excellence, new skills, cross-border and cross-sector mobility, innovation, and impact. Some of the individual fellowships on offer are directed early career postdoctoral researchers and offers them the opportunity to work on personal research projects by moving to a new country and even sector. The application is linked to a host institution.

The evaluation<sup>71</sup> of this large programme followed a mixed method approach. This comprised of an in-depth literature review that helped understand the nature and actions of those schemes. Large scale online surveys were conducted and were directed to funded researchers and a comparison group of other researchers together with funded organisations and a comparison group of organisations who applied but were unsuccessful. The comparison group for the researchers was created following a bibliometric analysis of data in Scopus that allowed to identify similarly successful, established researchers in similar research areas. The counterfactual group of organisations included only those who were above the quality threshold. That led to a sample of about 8,500 organisations and 11,000 researchers. The evaluation included “60 telephone interviews with EU stakeholders, national policymakers, research representatives and experts on human resources in research”. It also included 18 case studies of individual projects and a social network analysis of the programme.

18. **Emmy Noether Programme** (Hornbostel, et al., 2009). This is a programme with a multidisciplinary focus that run in Germany since 1999 and funded postdoctoral researchers. The goal of the programme was to support early career researchers to secure a permanent academic position early on, assist them in developing their independent research agenda, give them the means to establish international networks, promote women in academia, and prevent brain drain. The applicants had to already have two to four years of postdoctoral experience, a research stay abroad with international collaborations, and timely completion of research training.

The evaluation<sup>72</sup> of this programme answered whether the programme goals were achieved and if yes, if this was mainly due to the programme and if not if that was due to issues relating to the programme. The evaluation method focused only on physics and medicine awards and used a sample of 695 individuals comprising of awardees and unsuccessful applicants. They were asked to participate to an online survey which was then followed by interviews of a selected sample. The survey and interviews helped understand better the different experiences of the applicants and participants and shed light on the importance of the programme in their outcomes. To compare the professional outcomes of the two groups a bibliographic research collected information on publications (quantity of work) and citations (quality of work) of the sample covering the four years before the application and six years following the application. For the citations, the selection of papers made, allowed for a three-year window from the paper publication date for all individuals in the sample. They also collated online data on the professional positions of the people in the sample (academic versus non-academic, permanent, full-time) three years following the funding decision. The analysis of the data comprised of comparisons between the two groups.

19. **Erwin Schrödinger Fellowships with Return Phase** (Meyer & Bühner, 2014). The Erwin Schrödinger Fellowships with Return Phase is a Programme in Austria that supports Post-doctoral researchers across different disciplines who want to spend a

<sup>71</sup> <https://ec.europa.eu/research/mariecurieactions/sites/mariecurie2/files/interim-evaluation-msca-report.pdf>

<sup>72</sup> <https://link.springer.com/article/10.1007/s11192-009-0411-5>

period of 10 to 24 months in a University/Research Centre abroad. This scheme has been running since 1985 and by 2013 when the evaluation took place, it had 2,271 participants.

The evaluation<sup>73</sup> of this scheme followed a mixed methods approach including an online survey, a bibliometric analysis of the grant holders and an expert workshop. It looked at the impact of the programme on the researchers' output; their career development; their institutions and universities via transfer of knowledge and interconnectedness, and also the wider impact on the Austrian science system and the European Research Area. Even though the analysis of the survey was mainly descriptive, using a bibliometric analysis of the characteristics of the grant holders, a control group similar to the treated group was identified and their responses to the survey were also collected for comparison. Websites such as Scopus were used to identify the control group participants and characteristics such as age, gender, scientific discipline, publications and affiliation were used for the matching exercise. The final choice of the control group was random amongst individuals with the exact same characteristics as those in the treated group. In the end the treated group comprised of 703 respondents whilst the control group included 613 respondents.

The main outcomes included ordered replies on perceptions and also actual data on average publication count, average citation rate per publication, share of time spent on research given career phase, academic position, co-publication patterns and international mobility and affiliations. Even though this approach allowed for direct comparisons between scheme participants and non-participants, a more nuanced approach such as propensity score matching would have perhaps led to more robust matching. Furthermore, regression analysis would have offered a better understanding of the impacts of the scheme, net of the effects of other characteristics of the participants.

- 20. International Research Fellowship Programme** (National Science Foundation) (Martinez, et al., 2016). This US funded programme supports postdoctoral researchers in their early research careers for 9 to 24 months to undertake research outside of the USA with the aim of seeding productive international research collaborations. The evaluation used programme MI, and a survey of 1,039 applicants to the programme over a period from 1992 to 2009. It established counterfactual groups from the unfunded applicants using pre-award characteristics of applicants. This was undertaken in order to mitigate selection bias, and to try to ensure treatment and control groups were similarly motivated to engage in international collaboration. The evaluation used Propensity Score Matching to construct groups of 'statistically similar' groups of awardees and non-awardees. Comparisons were then made among similarly qualified groups. The evaluation found that fellows were more likely to have productive research collaborations with foreign researchers, and seeded collaborations that extended beyond the fellowship period.

---

73

<https://repository.fteval.at/86/1/Impact%20Evaluation%20of%20the%20Erwin%20Schr%C3%B6dinger%20Fellowships%20with%20Return%20Phase.pdf>

## 21. European Research Council Starting Grants (now Starting Independent

**Researcher Grants)** (Nedeva, et al., 2012). ERC Starting Grants are aimed at early career scientists ready to work independently to provide long-term funding to support excellent investigators and their teams to pursue ground-breaking, high quality/high risk research, which is expected to lead to advances at frontiers of knowledge. The grants are open to researchers of any nationality within 2 to 7 years of PhD completion (and have produced at least one important publication as main author) from any scientific field but they must conduct their research in an EU member state or associated country incl UK. The remit is broader than STEM to include physical sciences, life sciences and social sciences and humanities. Grants of up to €1.5m (plus €1m start-up costs) are awarded to hosts (which can include private organisations such as research labs) for a period of 5 years. Award rounds are held annually and it is anticipated that 413 Starting Grants will be made in 2021. Multi/interdisciplinary proposals are encouraged. Starting Grants therefore share many similarities with the FLF programme.

Review of outcomes of ERC Starting Grant applications in 2020<sup>74</sup> shows there were 3,272 applications for 436 awards (success rate of 13%): 124 awards (923 applications) in life science; 186 (1409) in physical sciences; and 126 (940) in social sciences and humanities. Across the awards, 62 were made to early career researchers in UK host institutions (2nd highest behind Germany), and 26 of the grantees were UK nationals (highest were German, Italian, French and Spanish). Overall, there were 273 (63%) males and 163 (37%) female grantees, and females were better represented in social sciences & humanities panels.

There are a number of evaluations<sup>75</sup> of European Research Council programmes including a synthesis study led by the University Manchester's Business School (Nedeva, et al., 2012<sup>76</sup>) which aimed to develop and apply a novel conceptual framework and methodology to measure, attribute and assess the impact and outcomes of ERC and its funding schemes (including Starting Grants). This study used a control group of researchers who passed quality thresholds but did not receive ERC grants to undertake a counterfactual impact evaluation.

Two methods were used to identify control groups of individuals: high scoring applicants, and analysis of a complex set of characteristics to identify matched pairs. To create matched pairs, three sets of attributes were used: demographics (age, gender, relationships and research field), approach (risk-taking, creativity disposition/activities), and standing (to measure resource conditions and to include organisational career, knowledge community position). This is referred to as the DAS framework. These

---

<sup>74</sup> [https://erc.europa.eu/sites/default/files/document/file/erc\\_2020\\_stg\\_statistics.pdf](https://erc.europa.eu/sites/default/files/document/file/erc_2020_stg_statistics.pdf)

<sup>75</sup> Others include: Investing in the European future we want (2017), Capturing career paths of ERC grantees and applicants: promoting sustainable excellence in research careers (2015), Qualitative evaluation of completed projects funded by ERC (2019)

<sup>76</sup> [https://erc.europa.eu/sites/default/files/document/file/eurecia\\_final\\_synthesis\\_report.pdf](https://erc.europa.eu/sites/default/files/document/file/eurecia_final_synthesis_report.pdf)

attributes were captured in a baseline survey of applicants and 11 variables<sup>77</sup> were used to create the match

The DAS framework was used to characterise the population and systematically measure differences between grantees and controls (at different time points and longitudinal comparisons within groups), and to identify 19 matching pairs for further analysis and allow for robust measurement of difference mediating for other differences unrelated to the grant. The grant is anticipated to impact on recipients' approach and their standing, and therefore for recipients to advance on these measures relative to the control group over time.

A variety of data collection methods were used in the evaluation:

- Descriptive survey of grantees and controls to measure impact on researcher (138/276 Starting Grant grantees and 46/105 controls, in the first cohort of applicants).
- Comparative case studies with grantees and controls (to identify and attribute causal changes in research and careers), sampled by type of programme and research field.
- Case studies of hosts (to identify changes in hosts or its sub-units, other organisational and funding landscape variables having an influence), sampled by number of grantees, type of organisation, level of specialisation, size, and international standing.
- Case studies of national research councils.
- Interviews with key policy stakeholders.
- Analysis of documents and individual-level bibliometrics.

Early impact on researchers measures included: reputation (from writing the proposal and getting the grant), start/maintain or expand a research group, ability to pursue their research agenda, less exposure to research politics, earlier/faster promotion and/or tenure, and possibly need to move hosts. Measures for impact on research content included: funding planned innovations (research findings affecting research practices of other researchers), funding to answer big questions, funding research that otherwise wouldn't have been funded/difficult to fund. Measures of impacts on researcher careers include: place on career ladder, relative autonomy, tenure, promotion, organisational mobility (but all of these are restricted by the system in place and starting position of the individual). Measures of impacts on hosts included: performance (internal funding, organisational decisions, career opportunities), organisational capabilities, investment in talent, and speed and scope of any changes (again these are affected by the starting position of the host, and other pressures which can make attribution difficult).

---

<sup>77</sup> Demographic (D1. Gender D2. Age bracket D4. Research domain); Approach (A2. Cognitive mobility A3. Perceived research novelty and risks); and Standing (Organisational career - S0. Researcher independence S2. Time for research S3. Long-term research direction independence; plus Knowledge community career - S7. Prestige of academic service, S9. Output productivity; and Research environments - S12. Perceived workplace reputation and performance).





## Appendix E: Potential administrative data sources

---

### Individual demographic data

#### ■ HESA Staff record (Higher Education Statistical Agency)

HESA collect data from HE providers about academic and non-academic staff who have a contract of employment with the HE provider. Data collected includes: personal and demographic characteristics, contracts and work patterns, jobs that academic staff had before entering higher education, and what they go on to do after leaving academia. HESA has collected data about HE in the UK since 1994 and has used consistent definitions within the staff survey since the 2017/18 academic year. The HESA staff survey contains metrics which capture R&I scale and scope.

These data present a complete record of staffing in HE and cover researchers (and identify early career researchers), it is captured annually, and has consistent variables. Each staff member has a unique identifier STAFFID which stays with them for the whole of their career in HE, enabling the identification and tracking of individuals (between HE providers). There are however still some issues with consistent use and reporting of original STAFFID when staff move between providers. Currently HESA does not require staff names but providers must keep a record of these in order to cross-reference to unique staff identifiers.

There was strong stakeholder support for this dataset, and it was regarded as useful for an evaluation of FLF. However, this will only cover those working in academia.

#### ■ Longitudinal Employment Outcomes (LEO)

LEO includes information from the Department for Education (DfE), Department for Work and Pensions (DWP), and HM Revenue and Customs (HMRC)

LEO is a linked administrative dataset created by combining data from the National Pupil Database (NPD), DfE; Higher Education Statistics Agency (HESA) data on students in HE, DfE; Individualised Learner Record data (ILR) on students in FE, DfE; employment data from the Real Time Information System (RTI) e.g. P45/P14 forms, and self-assessment tax returns, HMRC; and the National Benefit Database, Labour Market System and Juvos data, DWP. LEO experimental statistics were first published in 2016 and cover graduates from the 2003/04 academic year onwards.

Stakeholders tended to feel this data source would be less relevant to an evaluation of FLF as: FLF Fellows would be unlikely to be included in the graduating years currently covered; and the limited nature of outcomes captured would lend little to address the evaluation questions.

## Organisation level data

### ■ Inter-Departmental Business Register (IDBR)

The IDBR is a comprehensive list of information on around 2.7 million UK businesses in all sectors, estimated to cover 99 per cent of all businesses in the UK. The IDBR was introduced in 1994. The two main sources for the IDBR are Value Added Tax (VAT) and Pay As You Earn (PAYE) records from HMRC so it excludes small businesses unregistered for VAT and PAYE. Additional information comes from Companies House, Dun and Bradstreet and ONS business surveys, as well as the Department for Environment, Food and Rural Affairs (DEFRA) farms and the Department of Finance and Personnel, Northern Ireland (DFPNI). The IDBR contains a unique identifier (the enterprise reference number) that can be used to identify, track and link organisations.

Stakeholders felt this (and related) source could be useful as it covers information appropriate for an evaluation. However, there were concerns that the impact from Fellows may be 'lost in the noise'.

### ■ Business Population Estimates (BPE)

This draws on IDBR, Labour Force Survey and HMRC. The BPE provides the only official estimate of the total number of private sector businesses in the UK at the start of each year including employment and turnover: number of employees, legal status, industry and geography. It draws on the IDBR, Labour Force Survey and D HMRC Self-assessment data. It has existed in its current form since 2010.

Stakeholder feedback as above (see IDBR).

### ■ Business Structures Database (BSD)

The BSD is an annual extract of information from the IDBR. It began in 1997 and is available to access through the UK Data Service. The BSD contains information on employment (and number of employees), turnover, Standard Industrial Classification, legal status, foreign ownership, birth (company start date), death (termination date of trading) and various geographical variables.

The Enterprise Research Centre (ERC) has been formed to enable better exploitation of firm-level micro data, it has been working to integrate data from BSD with the Community Innovation Survey and Annual Business Survey and with the Employer Skills Survey. By linking these datasets the ERC can explore a wider variety of indices on firm characteristics and take longitudinal analysis to view growth

Stakeholder feedback as above (see IDBR). Innovate UK (2018) note how although this is a key source of data on the performance of businesses it has lags in uploading data and a lack of clear timestamps meaning it can be difficult to draw conclusions on cause and effect between R&I programmes and business outcomes. Also that R&D investment and employment of R&D professionals are not comprehensively covered.

## ■ Financial Analysis Made Easy (FAME)

FAME is a UK companies database and contains information for 3.8 million companies in the UK and Ireland. As well as for 4 million inactive companies available for historical research. For the top 1.5 million companies, available data includes: contact information, activity details, 29 profit and loss account and 63 balance sheet items, cash flow and ratios, credit score and rating, security and price information (listed companies only), details of holdings and subsidiaries, names of current and previous directors, heads of department, shareholders, news.

Stakeholder feedback as above (see IDBR).

## ■ HE Business and Community Interaction (HE-BCI) survey

The HE-BCI survey, run by HESA and now mandatory for all HE providers in England, Wales and Scotland, collects financial and output data related to knowledge exchange (KE) for each academic year since 1999. The survey includes information on business and public or third sector involvement in research, consultancy, and the commercialisation of intellectual property. The HE-BCI is the main mechanism for measuring the volume and direction of interactions between UK HE providers and business and the wider community. The survey collects information on the infrastructure, capacity and strategy of HE providers, and also numeric and financial data regarding third stream activity (that is activities concerned with the generation, use, application and exploitation of knowledge and other HE provider capabilities outside academic environments, these being distinct from the core activities of teaching and research). It collects information at the level of the HE provider rather than teams, functions or departments within it.

HE-BCI is highly valued. It is the key data source for the Knowledge Exchange Framework (KEF) and currently provides data for the majority of metrics used (additional data is provided by Innovate UK and Elsevier see below). HE-BCI is used to provide metrics on working with business, working with the public and third sector, skills, enterprise and entrepreneurship, local growth and regeneration, and IP and commercialisation. However it is only focused on HE providers. Other sources for information on entrepreneurship and spin-offs could include Companies House which would have broader coverage.

Stakeholder feedback again indicated a concern that the impact of FLF would be difficult to identify at the institutional level so felt HE-BCI may not be appropriate.

## ■ Longitudinal Small Business Survey (LSBS)

The LSBS is a large-scale telephone survey of small business owners and managers, commissioned by the Department for Business, Energy and Industrial Strategy (BEIS). This survey is the latest in a series of annual and biennial Small Business Surveys (SBS) dating back to 2003. The latest version covers the period from 2015 to 2019 and is available through the UK Data Service. The LSBS includes information on business performance and factors affecting this, including: employment and turnover, ambition and expectations of future performance, access to finance and use of business support.

Stakeholders felt this would be of limited relevance for the evaluation of FLF given the small numbers of SMEs involved in the programme.

### ■ **SME Finance Monitor**

The SME Finance Monitor provides survey data on access to finance amongst SMEs since 2011. Administered by BVA BDRC, a business insight consultancy, this survey includes over 160,000 interviews (4,500 each quarter), on SME sentiment around past and future finance needs. It explores demand for external funding among SMEs and the response to requests for funding made to banks in the last 12 months. The data are available through the UK Data Service.

Stakeholder feedback as above (see LSBS).

### ■ **HESA income data**

HESA has collected finance data from HE providers in Wales, Scotland and Northern Ireland covering income, expenditure, balance sheet, statement of gains and losses, capital expenditure and senior staff pay (where available). Key Financial Indicators are then derived from collected data. From 2018/19 the Office for Students (OfS) has been collecting financial data from HE providers in England, including Alternative Providers (APs) before this it was under HESA's remit. HESA still publishes data for English HEIs alongside those in the other UK nations.

Stakeholders have confidence in HESA data but as with other organisation level data have concerns over the extent to which FLF impact can be identified. In addition, these data only cover HE providers so do not cover the full remit of FLF influence.

### ■ **Community Innovation Survey (CIS)**

The Community Innovation Survey (CIS) based innovation statistics are part of the EU science and technology statistics. Surveys are carried out with two years' frequency by EU member states and a number of ESS member countries. Compiling CIS data is voluntary to the countries, which means that in different surveys years different countries are involved.

The CIS is a survey of innovation activity in enterprises. The harmonised survey is designed to provide information on the innovativeness of sectors by type of enterprises, on the different types of innovation and on various aspects of the development of an innovation, such as the objectives, the sources of information, the public funding, the innovation expenditures etc. The CIS provides statistics broken down by countries, type of innovators, economic activities and size classes.

New microdata release normally takes place two and half years after the end of the survey reference period. CIS microdata are available as scientific-use files (SUF - partially anonymised data) and as secure-use files in the Safe Centre (SC) at Eurostat's premises in Luxembourg.

## ■ Beauhurst

The beauhurst data is a comprehensive company level dataset which covers all incorporated UK businesses. It's scope is extensive in terms of the breadth and scale of the data it captures at the individual business level including ownership, corporate structure, accounting data, financing and capital raising, key employees etc.

## Researcher and research outputs level data

### ■ CIPHER.ai, patents data

CIPHER uses artificial intelligence and machine learning to classify patents according to customer specifications. The platform in its current form was invented in 2017. The platform can also aggregate, analyse and visualise data on the ownership and use of patents. The platform can classify 61 million patents in an hour as it uses machine learning.

Stakeholders noted this can be costly to access, cautioned that there will be a time-lag between patents being filed and published, and that patents have more relevance as an outcome measure to some disciplines than others (eg engineering and science) and to organisations with a greater patent budget. Also that a range of patent related data would be useful to collect including filing for patent, published patents and extending patents, as the longer that patents are supported indicates continued impact. Feedback indicated that other sources of patent data would be Espacenet (or esp@cenet a free online search service for patents and patent applications), the European Patent Office, the World Intellectual Property Organisation (WIPO) and UK Patents Office.

### ■ Research Excellence Framework (REF) submissions data

The REF started in 2014 and replaced the Research Assessment Exercise. HE providers submit data for all research staff to the relevant UK higher education funding body, e.g Research England. Research staff are assessed on: quality of outputs (e.g. publications, performances, and exhibitions), impact beyond academia, and the environment that supports research. The current REF 2021 had a submission date of 31st March 2021.

Submissions will be published on the REF website in Spring 2022 and will include a list of research groups, a list of submitted research outputs and accompanying data for each output, data on doctoral degrees awarded and research income, and submitted textual information about impact and the research environment plus submitted impact case studies (as a searchable database). However HEIs can exclude parts of their submissions from publication where publication is likely to cause harm to an individual or organisation perhaps due to commercial sensitive or protection of IP rights or premature release of information about public policy development

Stakeholder feedback indicated that REF submission data could be useful but some cautioned that these data could lack granularity, be very out of data (if using 2014 submissions), and potentially highly biased and unrepresentative.

## ■ Dimensions database

Dimensions is a bibliographic data source that indexes data using machine learning and cloud computing to enable tracking across the research cycle. Its basic search and browsing functions are free but advanced functions require payment (although this can be waived for non-commercial research projects). The dataset includes: funder, research organization, researcher, category status, publications, policy documents, and patents, and proprietary data from research organisations, publishers, or funders. Users can develop indicators exploring research income and output, citations and Altmetric attention, and collaboration activity and open access distribution.

Stakeholder feedback suggested that Dimensions could be useful however, some studies have indicated limitations with this (and related) source. A recent study (Guerrero-Bote et al, 2021 ) compared SCOPUS and Dimensions and found Dimensions' coverage was 25% greater, but Dimensions lacked affiliation data in relation to country and/or institution in almost half of its documents which can impede analysis. Another study that took place in 2019 (Martín-Martín et al, 2021 ) using a seed sample and analysis of citations compared six bibliographic databases including Dimensions, Web of Science and SCOPUS, to explore relative coverage differences across subjects. This found above 80% overlap between Dimensions and SCOPUS, and between Dimensions and Web of Science. It also found that Dimensions, SCOPUS and Web of Science tend to have similar coverage of each field (8 aggregated subject areas); and all have much better coverage of Chemical & Material Sciences and of Life Sciences & Earth Sciences; and much lower coverage of Business, Economics & Management, of Social Sciences, and particularly of Humanities, Literature & Arts. Although there is a high degree of overlap (particularly in STEM fields), the results of this study suggest that for identifying humanities and social sciences citations using a combination of sources may be advisable as there are many citations that one source finds that another does not, and perhaps Microsoft Academic could also be used to fill these coverage gaps.

## ■ Researchfish

Researchfish uses technology and algorithms to collect outcomes and outputs of research from the web, external data sources and researchers themselves including: publications, collaborations, further funding, next destination, engagement activities, influence on policy, research tools and methods, research databases and models, IP and licensing, patents, medical products and clinical trials, artistic and creative products, software and technical products, spin outs, awards and recognition, other outputs and knowledge and use of facilities and resources.

Researchfish is used widely by UKRI and its associated Research Councils. Other evaluation work has drawn on Researchfish including: ongoing evaluation of EPSRC Fellowships, evaluations of BBSRC Fellowships (David Phillips and Discovery Fellowships), and an evaluation of the Springboard programme funded by Wellcome Trust and others. Generally, these evaluations have focused on citation analysis and can be supplemented by other sources such as SCOPUS or discipline specific sources such as PubMed. Stakeholder feedback suggests this could be useful as it captures a



number of metrics of ‘success’ however as Researchfish data are self-reported and not mandatory this can lead to concerns over quality and completeness. They suggested UKRI data and data sources, such as the **Gateway to Research** Portal, could also be used to supplement and cross-referenced with Researchfish.

### ■ ResearchGate network

ResearchGate is a social networking site for researchers where members can upload research outputs including grey literature, connect with other researchers, view statistics for reading and citations of their work, collaborate by posting queries or project updates, and access the site’s job board. Starting in 2008 ResearchGate now has over 20 million members worldwide.

Stakeholder feedback suggests this might be a useful source. However many researchers do not have a profile on the site and it largely relies on individuals updating their profiles, so it will not have complete coverage.

### ■ Scopus

Scopus is owned by Elsevier and it is the world’s largest curated abstract and citation database with over 82 million items, it is source-neutral and is updated daily. It has an underlying metadata architecture to connect individuals, ideas and institutions. Scopus is a subscription-based database of peer-reviewed literature: scientific journals, books and conference proceedings, covering 25,000 active titles and 7,000 publishers. It covers science, technology, medicine, social sciences, and arts and humanities: 32% of content is social sciences focused, 27% physical sciences, 25% health sciences and 16% life sciences. It includes tools to track, analyse and visualise research including the SciVal system which enables research activity and performance to be systematically evaluated at institution or individual researcher level, and Author Evaluator to analyse an individual researcher’s publishing output and research impact. It includes metrics on awards volumes, collaboration, publications, views of output, citations and economic and societal impact. Societal impact includes media exposure with a weighted measure based on subject.

This is a key data source used in evaluations. Scopus provides data used in the co-authorship metric used in the Knowledge Exchange Framework (KEF), where outputs are collected for each HE institution and are analysed for the presence of non-academic authors and then the proportion of outputs with non-academic authorship is derived. It has also been used in other evaluations of fellowship programmes including ongoing evaluation of EPSRC Fellowships, and evaluations of Marie Skłodowska-Curie Actions schemes and of Erwin Schrödinger Fellowships (the latter two used SCOPUS data to identify a control group rather than to measure outcomes).

However, research indicates that documents are included from a pre-selected list of publications, SCOPUS does not cover non-peer-reviewed scientific documents, it has lower coverage in Humanities and Social Sciences than other sources, it is slower at indexing, and is not free to use (Martín-Martín et al, 2021).



### ■ **Web of Science/Data Citation Index**

The Web of Science is a subscription-based citation index drawing upon databases across Science, Social Sciences and the Arts and Humanities including books and conference proceedings. The Data Citation Index catalogues research data deposited at repositories around the world and across disciplines, and links this with research literature in the Web of Science. Web of Science claims to be the most trusted citation index for scientific and scholarly research. The collection contains over 21,000 peer reviewed, high-quality scholarly journals published worldwide (including Open Access journals) in over 250 science, social sciences and humanities disciplines. It was one the first record of published research/bibliographic sources available and combined the Science Citation Index, Social Sciences Citation Index and the Arts & Humanities Citation Index.

However other newer sources have emerged that provide metadata on scientific documents and citation links including SCOPUS and Google Scholar (the latter is free to access and has an inclusive approach to indexing materials and is argued to have more comprehensive coverage). See also criticisms of SCOPUS and Dimensions above.

### ■ **Individual HE provider Online Research Archives**

HE provider online archives contain research produced by authors or units based at that provider and not controlled by commercial publishers. These may include doctoral theses, masters dissertations, project reports, briefing papers and out-of-print materials. The coverage differs by provider and stakeholders felt these may not provide much in additional material.

### ■ **Google Scholar**

Google Scholar provides a simple way to broadly search for scholarly literature. It allows for searching across many disciplines and sources for: articles, theses, books, abstracts and court opinions; and from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar focuses on scholarly research, so is likely to be biased towards academic research outputs.

Google Scholar Metrics show the visibility and influence of recent articles in scholarly publications, and summarises recent citations to many publications. It includes metrics such as h-index and h-median metrics. It is organised into research categories (broad areas of research and then further subcategories). As noted above Google Scholar is free to access and has an inclusive approach to indexing materials and is argued to have more comprehensive coverage than older abstract and citation indexes.

### ■ **RePec (Research Publications Economics)**

RePEc (Research Papers in Economics) is a database culminating from a collaborative project across over 100 countries to enhance the dissemination of research in Economics and related sciences. It is an example of a discipline specific data source and there are likely to be others that will be useful for the FLF evaluation.

It is a decentralized bibliographic database (pulled together from many sources) of working papers, journal articles, books, books chapters and software components, all maintained by volunteers. The collected data are then used in various services that serve the collected metadata to users or enhance it. To date, has drawn from over 2,000 archives to amass over 3.5 million research items from 3,600 journals and 5,300 working paper series; and over 62,000 authors are registered. It is free to use but given its decentralised nature can be difficult to search and some data is available on request only.

#### ■ **Altmetric.com**

Altmetrics are an evolving field and attempt to provide an alternative to traditional journal impact indicators and to capture and measure the diversity of communication methods. A leader in the field is Altmetric.com. This monitors mentions of published research online and has done so since 2012. Currently the Altmetric database contains 166.9 million mentions of over 32.1 million research outputs. Sources include: public policy documents, mainstream media, online reference managers, post-publication peer-review platforms, Wikipedia, Open Syllabus Project, patents, blogs, citations, research highlights, social Media (eg Twitter), and multimedia and other online platforms. Altmetric provides quantifiable metrics as well as qualitative information. The platform uses identifiers such as PubMedID, arXiv ID, ADS ID, SSRN ID, RePEC ID, URNs, ISBNs and DOIs.

The UK CDR (UK Collaborative on Development Research) note how altmetrics may not be very useful in quantifying impact as they have 'very short-half lives – and so could be registering hype/buzz as opposed to more subtle longer term change'. Stakeholders also had reservations about the utility of altmetrics in terms of the 'real-world research characteristics' captured, what it added to other sources, and concerns about the cleanliness of the data.