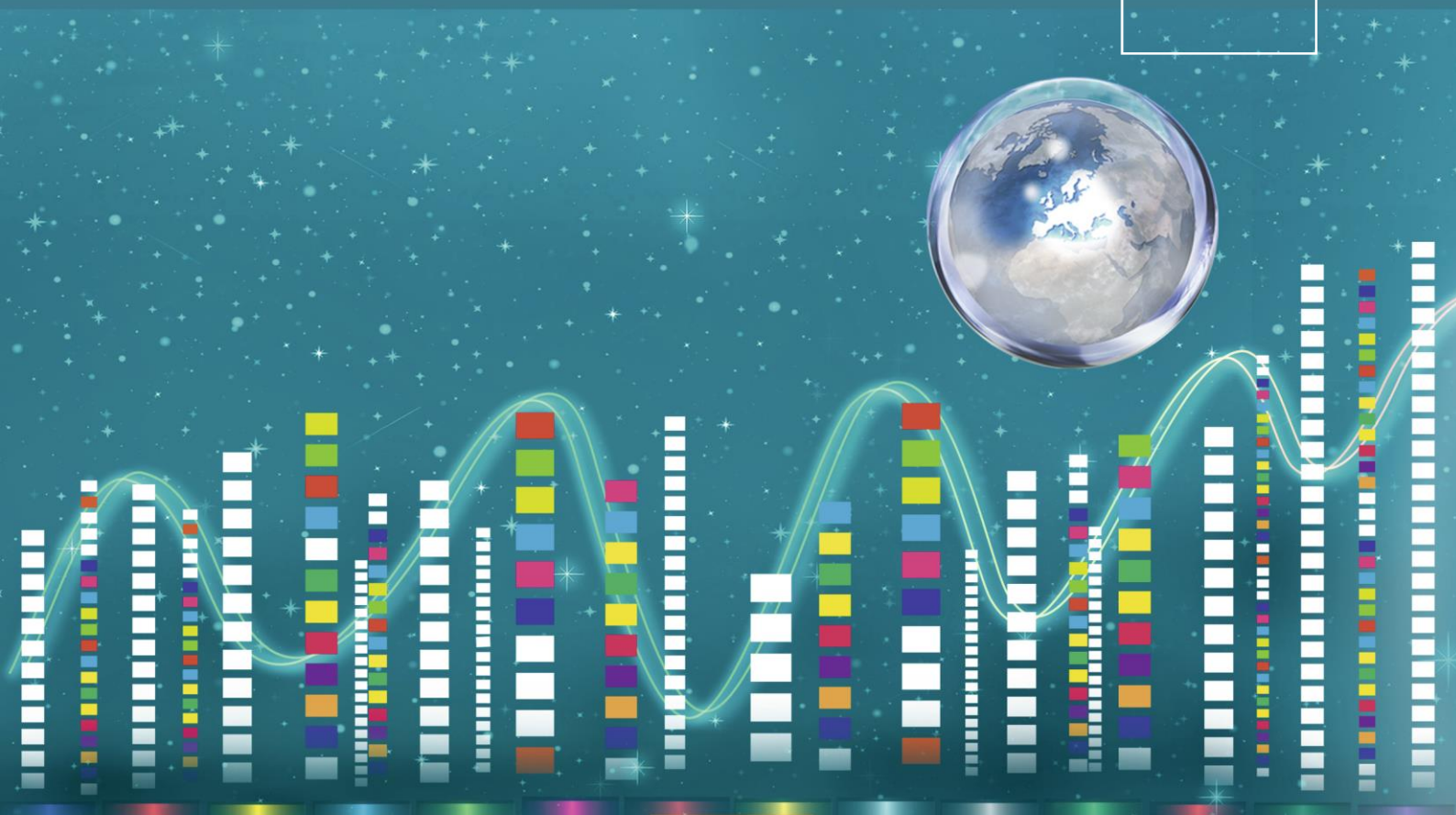
European Commission

Study on the proposal evaluation system for the EU R&I framework programme

Final report

Independent Expert Report



Research and Innovation

Study on the proposal evaluation system for the EU R&I framework programme

European Commission
Directorate-General for Research and Innovation
Directorate G — Common Policy Center
Unit G2 — Common Programme Analysis & Regulatory Reform
Contact Diletta Zonta
Email diletta.zonta@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in October 2021

PDF	ISBN 978-92-76-42545-8	doi: 10.2777/16211	KI-01-21-369-EN-N

Luxembourg: Publications Office of the European Union, 2022

EUROPEAN COMMISSION

Study on the proposal evaluation system for the EU R&I framework programme

Final report

Daniela Rodriguez-Rincon, Carolina Feijao, Cagla Stevenson, Hamish Evans, Alexandra Sinclair, Simon Thomson and Susan Guthrie

2022 Directorate-General for Research and Innovation EN

PREFACE

RAND Europe was commissioned by the Directorate-General for Research and Innovation (known by the acronym DG RTD) to conduct a comprehensive review of proposal evaluation processes used in Horizon 2020 (H2020), and to conduct international benchmarking of these processes against accepted practice and the wider knowledge base in this area. This report sets out the findings of that study and is intended to inform the design of proposal evaluation processes for Horizon Europe. The work may also be of interest to other research and innovation (R&I) funding organisations and the wider European research and innovation community. This report constitutes the main findings of the analysis. It is accompanied by an annex document, which provides more detail on the methods and the data used and produced for reference, as well as the specific analyses conducted.

This study incorporates bibliometric analysis conducted by Clarivate and expert advice from Prof. Jonathan Grant, Prof. Liv Langfeldt, Prof. Ana Marušić and Prof. Ulf Sandström.

CONTENTS

ABBREVIATIONS AND ACRONYMS

CFI	Canadian Foundation for Innovation
CIHR	Canadian Institutes of Health Research
CNCI	category-normalised citation impact
COVID-19	Coronavirus 19 disease
CSA	Coordination and Support Actions
DG RTD	Directorate-General for Research and Innovation
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
EIC	European Innovation Council
EIT	European Institute of Innovation and Technology
EJP	European Joint Programme
ERC	European Research Council
ESR	evaluation summary report
EU	European Union
H2020	Horizon 2020
IA	Innovation Actions (Horizon 2020)
IER	individual evaluation report
MSCA	Marie Skłodowska-Curie Actions
MSCA IF	Marie Skłodowska-Curie Actions Individual Fellowships
NCN	National Science Centre [Poland]
NHMRC	National Health and Medical Research Council [Australia]
NRF	National Research Foundation [Singapore]
NSF	National Science Foundation [United States]
OR	odds ratio
PCP	Pre-commercial Procurement
PoC	Proof of Concept
PPI	Public Procurement of Innovative Solutions
R&I	research and innovation
RAG	red, amber, green
RIA	Research and Innovation Actions (Horizon 2020)

SME	small to medium-sized enterprise
SNSF	Swiss National Science Foundation (in German: Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung)
UKRI	United Kingdom Research and Innovation
ZonMw	the Dutch organisation for health research and development

ABSTRACT

This comprehensive review of proposal evaluation processes in the European Union's research and innovation framework programmes is intended to inform the development of proposal evaluation processes in Horizon Europe. Based on a review of the literature, case studies, and an analysis of existing and new data sets on proposal evaluation in Horizon 2020, we find that, broadly speaking, proposal evaluation processes were fair and transparent. However, there may be scope to improve consistency and the feedback provided, and to reduce the burden and conservatism of the process. Based on these findings, we identify 10 recommendations for Horizon Europe:

1. Conduct regular reviews of the fairness of the process.
2. Encourage more female applicants.
3. Limit the use of multi-stage processes.
4. Explore and experiment with novel approaches, such as double-blind review and lotteries.
5. Streamline the use of consensus meetings.
6. Improve the clarity of assessment criteria and ensure that the evaluation process is centred around them.
7. Consider more targeted use of reviewers depending on their expertise.
8. Explore the scope for use of novel technologies to improve the process.
9. Continue to provide constructive feedback, especially for unsuccessful applicants.
10. Consider routes to ensure that innovative R&I is supported, such as targeted funding streams.

CHAPTER 1: INTRODUCTION

1 Context and purpose

Understanding and evaluating the effectiveness and efficiency of proposal evaluation systems is crucial to the effective delivery of research and innovation. Public R&I funding is central to the research system because it helps develop the R&I workforce, creates infrastructure for doing research and enables the conduct of research studies. Research funding is inextricably linked to the values, culture and conduct of the R&I ecosystem, and its processes should be consistent with the values and norms of that wider system (Adam et al., 2018). Its links to key actors in academia are particularly important, since research funding can also create important levers that shape the culture and behaviour of researchers (Marques et al., 2017; Grant, 2021). Proposal evaluation processes based on peer review are well established and widely considered as a respected approach to funding allocation. Although support for peer review–based proposal evaluation processes has traditionally been strong among academics and researchers (Wooding and Grant, 2003), over recent decades, criticisms have emerged, both within and beyond the research community. Issues highlighted by critics of peer review include waste and burden, bias against some types of research or researchers, concerns that peer review may be risk averse, and – perhaps most significantly – failure to effectively safeguard the quality and integrity of research (Fang, Bowen, and Casadevall, 2016). With this in mind, it is essential to assess proposal evaluation processes to ensure that grant evaluation systems are achieving their desired aims, enabling them to support the best R&I.

The purpose of this study is to provide a comprehensive review of proposal evaluation processes in EU R&I framework programmes to inform the development and, where required, adaptation of the approach taken to proposal evaluation in Horizon Europe. This work is timely because Horizon Europe is due to be launched in 2021. We have been commissioned by DG RTD to conduct this work, which consists of a comprehensive review of the processes used in H2020 and international benchmarking of these processes against accepted practice and the wider knowledge base in this area. The research questions for this work are the following.

1. How relevant[1] is the current proposal evaluation system to achieve the objectives of H2020?

2. How efficient[2] is the current proposal evaluation system?

3. How effective[3] is the current proposal evaluation system?

4. What are the lessons learnt and areas for improvement for the future? What needs to be adapted to accommodate the novel features of Horizon Europe?

---

[1] We consider relevance of the proposal evaluation system to be the extent to which it is efficient and effective and thus allows H2020 to deliver on its objectives: Is it an appropriate system for the needs of a modern R&I funder of the scale and scope of the framework programmes?

[2] We consider efficiency of the process to be the burden of the process on different individuals involved in submitting and evaluating proposals.

[3] We consider effectiveness of the process to be the extent to which it delivers a good outcome, taking into account some of the known critiques of peer review–based processes: Does it deliver a fair, consistent appropriate assessment of the quality of applications?

Combining the research questions with our existing knowledge of the evidence around the effectiveness and efficiency of proposal evaluation systems, we developed an initial conceptual framework for the study (Figure 1).

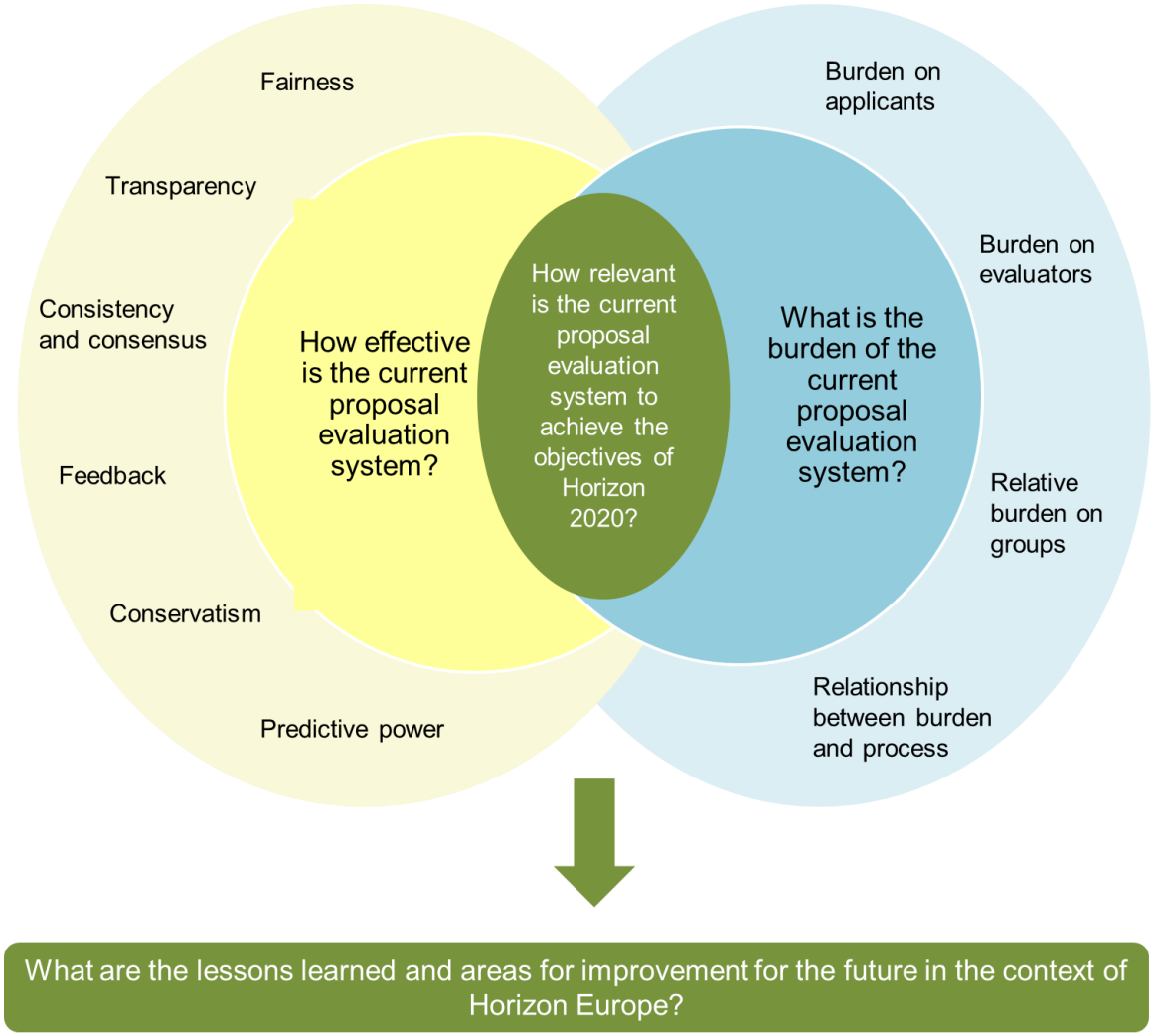


Figure 1: Conceptual framework

2 Approach

To implement the study, our conceptual framework is operationalised across four primary tasks. Our overall approach delivers cross-cutting insights spanning the key study questions as set out in Table 1, which characterises the key aims and outputs of the different tasks and sub-tasks.

Table 1: Study approach		
Task and method	Key questions addressed	Brief summary of tasks
Task 1: Literature review	What is the existing evidence regarding the characteristics of proposal evaluation systems that influence their effectiveness and burden? What novel approaches have been used to address these issues, and do they work?	We conducted a rapid evidence assessment of literature on the effectiveness and burden of proposal evaluation processes covering literature in English from the past 10 years.
Task 2: Review of H2020 process – Analysis of scoring data	Is the proposal evaluation process fair and replicable? How do scores and outcomes vary by personal and institutional characteristics of applicants (if at all)? How do scores and outcomes vary depending on the reviewer (if at all)? How do these intersect?	DG RTD provided us with data on 160 858 applications to H2020. This included individual scores on applications provided by reviewers and aggregate scores, as well as information on lead applicants and reviewers, such as gender, organisation type and nationality. We analysed the data to look for any evidence of bias in evaluation processes and to assess replicability and consistency among reviewers.
Task 2: Review of H2020 process – Analysis of reviewers survey data	What are the perceptions of reviewers on the effectiveness of different elements of the proposal evaluation processes? What are areas of concern, strengths and weaknesses?	DG RTD provided us with data from 3 602 respondents to a survey of reviewers for 2016/2017 H2020 processes, which gave information on their views on different aspects of the proposal evaluation process. We analysed the qualitative comments and quantitative scores provided.
Task 2: Review of H2020 process – Bibliometric analysis	Is the proposal evaluation process effective in predicting outcomes?	We conducted a bibliometric analysis of a sample of 3 815 applicants to ERC and MSCA[4] from 2014–2016, looking at their performance prior to and following application, assessing differences between successful and unsuccessful applicants and generating breakdowns by gender and country of applicant.
Task 2: Review of H2020 process – Survey of a sample of applicants and reviewers	What is the burden of the proposal evaluation process, and how is it distributed across groups and across stages of the process and different actions? Is the burden proportionate to the outcomes?	We conducted a survey of applicants and reviewers from the past two years, asking them to provide quantitative assessment about the amount of time they spent on different aspects of the proposal evaluation process. We

[4] We identified a sample of 5 004 applications to ERC/MSCA IF including a roughly equal mix of successful and unsuccessful applicants. We used information on name and email address of applicants to match these applicants in the Web of Science database and identify relevant publications. We were able to match 76.2% of applications, resulting in a sample for analysis of 3 815 applicants.

		received responses from 5 369 applicants and 7 049 reviewers.
Task 2: Review of H2020 process – Transparency analysis of documentation	How transparent is the proposal evaluation process across schemes?	Through desk research, we mapped the extent to which information is available on different aspects of H2020 processes, using a transparency framework, and we compared this to international examples.
Task 2: Review of H2020 process – Analysis of feedback letters	What is the nature and quality of feedback provided to applicants to H2020 funding schemes? How does it vary?	We analysed a sample of 300 feedback letters across actions, for both successful and unsuccessful applicants, coding characteristics of the feedback provided thematically.
Task 2: Review of H2020 process – Desk research	What is the structure of the proposal evaluation processes in H2020, and how do they vary?	We reviewed documentation on H2020 proposal evaluation processes to understand how they work, as well as on upcoming plans for Horizon Europe.
Task 2: Review of H2020 process – Interviews with European Commission staff	What is the rationale for the design of the H2020 evaluation process, and how does it align with the aims of H2020 and specific actions?	We conducted interviews with eight Commission staff responsible for the operation and oversight of H2020 proposal evaluation processes.
Task 3: Case studies	How do other funders structure their proposal evaluation processes? What is the rationale for their approach, and what measures, if any, have they taken to address effectiveness and burden? What evidence, if any, is there of the effectiveness of these approaches?	We conducted case studies of seven international R&I funders[5], through desk research and interviews.
Task 4: Analysis and lessons learnt	Across tasks, what are the lessons learnt about the proposal evaluation process, and how can these inform Horizon Europe?	We used a framework synthesis approach to map evidence from the different tasks against our conceptual framework and to identify key findings and recommendations.

There are a number of caveats to our findings that are important for the reader to bear in mind when reviewing the evidence provided:

1. Burden analysis: Our analysis of the burden of H2020 proposal evaluation processes is based on a survey of applicants and reviewers over the past 2 years. Collecting appropriate retrospective data on the time burden associated with tasks is challenging because survey respondents may not recall the specific amount of time spent on detailed tasks. By pooling

[5] The case studies looked at the Canadian Foundation for Innovation (CFI); the German Research Foundation (known by its German acronym, DFG); the National Health and Medical Research Council (NHMRC) (Australia); the National Research Foundation (NRF) (Singapore); the National Science Centre (known by its Polish acronym, NCN) (Poland); the Swiss National Science Foundation (SNSF); and United Kingdom Research and Innovation (UKRI).

responses and analysing the median response across task and stakeholder group, we aimed to minimise effects of respondent's inaccurate recall.

2. Analysis of evaluators survey data: Analysis is based on an existing data set provided from a survey conducted in relation to reviewers participating in H2020 in 2016/2017. These data are a few years out of date, and processes may have changed in the intervening period. The survey provides a picture of the views, perceptions and experiences of reviewers. Therefore there may differences of opinion and biases, and the results should be considered as the views and perceptions of this group, rather than the absolute truth. Views of others – for example, applicants – are not reflected in this data set. Finally, evaluators cover a range of different H2020 programmes and therefore may have different experiences and give different feedback based on what programme they were involved with.

3. Analysis of feedback letters: We conducted an analysis of a sample of 246 feedback letters provided to applicants across six H2020 programmes, comprising a mix of successful and unsuccessful proposals. This is a reasonable sample, but it is still small relative to the overall number of such letters produced over the duration of H2020. Therefore, some caution must be taken in generalising these observations to the programme as a whole.

4. Case studies: We have conducted seven case studies reviewing international practice in proposal evaluation. These provide interesting examples of practice that may offer learning for Horizon Europe. However, we note that practice at these different funding organisations has typically not been evaluated for its effectiveness, so we cannot assume that common practices are necessarily effective in delivering desired outcomes.

5. Bibliometric analysis: We have used bibliometric analysis to review the predictive power of proposal evaluation processes (by looking at post-award performance of both successful and unsuccessful applicants) and also to assess the bibliometric performance associated with a successful application by gender and country. We note that this provides a limited picture of the range and nature of the potential achievements and outputs of R&I activities. Reflecting the scope and nature of the different actions, and the requirements of the analysis, this analysis has also been limited to a total sample of around 5 000 ERC and MSCA IF applicants.

Despite these limitations, we are confident in the robustness of our findings, which build on triangulation and synthesis of evidence from multiple methods and sources. Please note that all differences reported among groups within this document are statistically significant.

3 Structure of the report

This document is the final study deliverable, *Final Report D4*, and as such provides an overview of key findings across the entire project. This report is supported by a supplementary annex document that provides detailed methods, data and results for each component of the work. The next chapter provides an overview of H2020 proposal evaluation processes. Chapter 3 then provides an overview of key findings from our analysis. Finally, Chapter 4 sets out recommendations for Horizon Europe.

CHAPTER 2: INTRODUCTION TO PROPOSAL EVALUATION PROCESSES IN HORIZON 2020

H2020 is the eighth framework programme of the European Union and was the main mechanism for funding research and innovation in Europe for the period 2014–2020. With a budget of nearly EUR 77 billion, H2020 was larger than any previous EU R&I programme. As well as supporting R&I within Member States, H2020 was also open to 'associated countries', which are affiliated with the programme at a national level[6], and to 'third countries'[7]. The primary goals of H2020 are to ensure that Europe produces world-class science, to remove barriers to innovation and to make it easier for the public and private sectors to work together. Funding through H2020 is set out in a number of multi-year work programmes, organised under three pillars, or priorities: Excellent Science, Industrial Leadership, and Societal Challenges. Along with the work programmes set out under these three pillars, there are also cross-cutting work programmes in H2020, including: Science with and for Society; Spreading Excellence and Widening Participation; the European Institute for Innovation and Technology (EIT); Euratom (a complementary funding programme to support nuclear research and training); and a pilot of the European Innovation Council (EIC), which will be rolled out on a wider scale during the next European framework programme.

Calls under H2020 represent different types of actions, or funding schemes. Depending on the type of action, the call will diverge in terms of the scope of what is funded, the reimbursement rate and the standard forms that are used, such as the lump sums provided for each cost. Additionally, the type of action also determines the specific criteria that will be used to evaluate proposals submitted under each call. The main types of actions are Research and Innovation Actions (RIA), which support basic and applied research to produce new knowledge and/or to explore the feasibility of a new or improved innovation; Innovation Actions (IA), which support plans for new, altered or improved products; and Coordination and Support Actions (CSA), which assist in processes supplemental to R&I activities, such as standardisation efforts, awareness raising, dissemination, communication, networking and coordination. Table 2 shows the key actions under H2020, as well as a brief description of their objective.

[6] EU-associated countries for H2020 are Iceland, Norway, Albania, Bosnia and Herzegovina, North Macedonia, Montenegro, Serbia, Turkey, Israel, Moldova, Switzerland, Faroe Islands, Ukraine, Tunisia, Georgia and Armenia.

[7] The term third countries refers to countries that are not members of the EU and countries or territories whose citizens do not enjoy the EU right to free movement. The most recent list of eligible countries can be found in *Horizon 2020 – Work Programme 2018–2020* (https://ec.europa.eu/research/participants/data/ref/h2020/other/wp/2018-2020/annexes/h2020-wp1820-annex-a-countries-rules_en.pdf).

Table 2: Main actions of H2020	
Type of action	Objective
Research and Innovation Actions	RIA comprise activities that aim to produce new knowledge and/or to explore the feasibility of a new or improved technology, product, process, service or solution.
Innovation Actions	IA comprise activities that aim to produce plans and arrangements or designs for new, altered or improved products, processes or services. IAs are used to turn existing scientific and technological insights into applications.
Coordination and Support Actions	CSA support standardisation, dissemination, awareness raising and communication, networking, coordination, support services, learning exercises, and dialogues that enable the development of new infrastructure.
Marie Skłodowska-Curie Actions	MSCA provide grants for researchers at all stages of their career – be they doctoral candidates or highly experienced researchers – and encourage transnational, intersectoral and interdisciplinary mobility.
European Research Council	The ERC funds projects based on scientific excellence in any field of research, conducted by a research team led by a principal investigator.
SME instrument	The SME instrument aims to support all types of innovative small to medium-sized enterprises with a strong ambition to grow internationally.

1 H2020 proposal evaluation processes

H2020 has issued more than 3 850 calls during its implementation[8]. Each call receives, on average, more than 1 500 proposals, although this varies significantly among calls and actions. Two main elements shape the awarding process of H2020 funding: proposal submission and proposal evaluation.

1.1 Proposal submission

Depending on the call description in the H2020 work programmes, applicants are invited to submit their proposals either in one stage (single-stage call) or in two stages (two-stage calls). Proposals are submitted according to the relevant proposal template. In the one-stage submission, applicants submit the full proposal, which will go through the full evaluation process. In a two-stage submission, there are two deadlines per proposal. In the first stage, applicants submit a short proposal for evaluation, consisting of two of the five sections of a full proposal, focusing specifically on the excellence of the application and its likely impact. Stage-one submissions are then assessed by external experts, and some participants are invited to submit a full proposal for stage two. The value of proposals

[8] According to data available on the Single Electronic Data Interchange Area (SEDIA) for funding and tender opportunities for H2020, accessed 12 October 2020.

admitted to stage two should be three times the available budget, and in any case, not less than two and a half times the available budget. Because of potential changes in the proposal between stages, experts once again review the proposal in terms of the criteria 'Excellence' and 'Impact', in addition to the new criterion of 'Quality and efficiency of implementation'.

1.2 Proposal evaluation

Proposal evaluation typically consists of three steps:

1. **Individual evaluation:** All proposals submitted under H2020 are evaluated by independent experts. More than 22 000 different independent experts, from 114 countries, have evaluated H2020 proposals so far. In total, more than 850 000 evaluations have been conducted. Experts are recruited through a continuously open call for expression of interest published in the Funding & Tenders Portal. Registration is open to experts of any nationality, including those outside of the EU and EU-associated countries. Experts are compensated for their time at a rate of EUR 450 per day.

A minimum of three external experts evaluate each proposal individually and remotely according to three common criteria: 'Excellence', 'Impact', and 'Quality and efficiency of the implementation'. Different actions will have different descriptions of these criteria depending on the objectives of the programme. For most schemes, reviewers assess each of these criteria on a five-point scale (as follows), and they provide comments on each criterion.

- 0: The proposal fails to address the criterion or cannot be assessed due to missing or incomplete information.
- 1 Poor: The criterion is inadequately addressed or there are serious inherent weaknesses.
- 2 Fair: The proposal broadly addresses the criterion, but there are significant weaknesses.
- 3 Good: The proposal addresses the criterion well, but a number of shortcomings are present.
- 4 Very good: The proposal addresses the criterion very well, but a small number of shortcomings are present.
- 5 Excellent: The proposal successfully addresses all relevant aspects of the criterion. Any shortcomings are minor.

In order to be recommended for funding, proposals must score above a certain threshold for each individual criterion, as well as meet an overall threshold for all criteria collectively. These thresholds vary among calls. For example, for the SME Concept and Feasibility Assessment awards, the score on each individual criterion must be at least 4, and the overall score, applying to the sum of the three individual scores, must be at least 13. On average, around 48% of applications meet the threshold; the proportion varies depending on the call.

2. **Consensus group discussion:** After proposals have been evaluated independently, reviewers convene in the consensus meeting. The consensus meeting is chaired by a rapporteur or a moderator from the European Commission or the Executive Agency and involves all the individual reviewers of the proposals. The rapporteur can also be an external expert not involved in the evaluation of the proposals. The meeting aims to find an agreement on the final comments and scores of the proposals. Meetings are

held on-site (in Brussels) or remotely (through teleconferences or through a comment box functionality in the evaluation tool)[9].

The rapporteur drafts the consensus report, which is an internal document. This report summarises the justification of scores, records any dissenting views, and records the consensus view that has emerged from the group's discussion.

3. **Panel review:** A panel review ensures consistency in the comments and scores given at the consensus stage for all the submitted proposals to the call. It recommends a list of proposals for funding in priority order, based on the agreed final scores. On average, around 13% of applications are recommended for funding. A panel review consists of a subset of experts or external experts not involved in the evaluations of the proposals. Panel review leads to the evaluation summary report, or ESR, which summarises the evaluation results and includes scores and supporting comments, and which is shared with the applicants.

There are specific adjustments to this broad process for different actions and calls. For example, some calls include an interview stage (e.g. commonly for the SME instrument), while others use a single-stage submission process but a two-step evaluation process, where an initial assessment is made based on some aspects of the application before the full applications is reviewed (this is typical for ERC processes). Also, the use and weighting of the evaluation criteria may vary among actions. A more detailed mapping of the proposal evaluation processes in H2020 and the nuances among actions is provided in Chapter 3 of the accompanying annex document.

[9] This is a functionality that enables reviewers to share comments and discuss the application and scores asynchronously.

CHAPTER 3: KEY FINDINGS

In this chapter, we set out the key findings of our analysis across tasks, drawing on all the evidence collected. More detail on the specific tasks and their individual findings is available in the accompanying annex document.

1 H2020 processes are broadly fair, but fairness should be monitored given the wider evidence of the risk of bias in proposal evaluation processes

Our analysis suggests that the process of proposal evaluation in H2020 is largely fair. For most characteristics – particularly gender and EU/non-EU nationality – we find that the chances of success are similar, suggesting a broadly fair process. However, after analysing mono-beneficiary programmes in more detail, we can identify some differences in outcomes among groups (Table 3, Table 4 and Table 5). Across all three actions, we also note that applicants from widening participation group countries had lower success rates than applicants from other EU Member States. For MSCA IF and ERC actions, applicants from EU-associated and third countries have significantly lower odds of success in comparison with applicants from the EU (0.33 and 0.28 lower odds, respectively). We also note a few other differences based on applicant affiliation, with research organisations (14.74%), public bodies (21.11%), and private for-profit organisations (8.85%) having the highest success rate for ERC, MSCA IF and the SME instrument, respectively.

For the SME instrument specifically, male lead applicants have 1.13 higher odds of success in comparison with female lead applicants, and applicants in non-private organisations[10] have 0.5 lower odds of success than applicants in private organisations. The latter we might expect given the nature of the instrument (which aims to support SMEs), but the differences in success rate by gender of lead applicant may merit further investigation. These differences by country and gender could reflect biases in the system, but they could also reflect differences in the quality of applications from these different groups, as we do not have an independent measure of application quality

[10] These are research organisations, higher or secondary education establishments, public bodies, or other organisations.

Table 3: Success rate by applicant gender (mono-beneficiary actions only)[11]			
Action	Gender	Success rate (%)	Number of applicants
MSCA IF	Female	15.84	26 333
MSCA IF	Male	14.17	38 185
ERC	Female	12.58	16 438
ERC	Male	12.94	40 747
SME instrument	Female	8.02	11 079
SME instrument	Male	8.93	53 753

[11] Note that overall success rates for MSCA IF here are higher than in official statistics due to limits within the dataset that mean that some unsuccessful awards cannot be matched to data on applicant characteristics. Therefore, this provides a useful comparison of relative success rates between groups but is not an accurate representation of overall success rates for this funding programme. Also note that ERC grants are excluded from this analysis, as there was no gender data available for these grants.

Table 4: Success rate by applicant nationality (mono-beneficiary grants only)			
Action	Nationality	Success rate (%)	Number of applicants
ERC	EU Member State	12.71	39 683
ERC	Widening participation group country	10.37	3 847
ERC	Other Member State	12.96	35 836
ERC	EU-associated country	15.66	4 860
ERC	Third country	12.15	12 642
MSCA IF	EU Member State	16.96	33 780
MSCA IF	Widening participation group country	14.11	5 437
MSCA IF	Other Member State	17.51	28 343
MSCA IF	EU-associated country	11.13	4 509
MSCA IF	Third country	12.77	26 230
SME instrument	EU Member State	8.71	48 886
SME instrument	Widening participation group country	5.00	10 140
SME instrument	Other Member State	9.69	38 746
SME instrument	EU-associated country	10.27	8 251
SME instrument	Third country	7.43	7 922

Table 5: Success rate by applicant affiliation (mono-beneficiary only)			
Action	Organisation type	Success rate (%)	Number of applicants
ERC	Higher or secondary education establishments	12.39	43 997
ERC	Research organisations	14.74	12 057
ERC	Private for-profit entities	11.60	854
ERC	Public bodies	4.71	191
ERC	Other	4.65	86
MSCA IF	Higher or secondary education establishments	14.41	49 152
MSCA IF	Research organisations	16.41	12 701
MSCA IF	Private for-profit entities	15.18	2 240
MSCA IF	Public bodies	21.11	251
MSCA IF	Other	13.14	175
SME instrument	Higher or secondary education establishments	5.3	509
SME instrument	Research organisations	n/a	0
SME instrument	Private for-profit entities	8.85	63 467
SME instrument	Public bodies	5.9	407
SME instrument	Other	3.99	676

To investigate this further, we analysed whether any characteristics of the reviewer influenced the outcomes for different groups. We found that in most cases characteristics of reviewers did not affect outcomes for different groups of applicants. However, we did identify two cases where the reviewer characteristics did affect the likelihood of success for one group. First, for MSCA IF, we find that proposals submitted by male applicants have lower odds of succeeding when assessed by male reviewers (Table 6). Second, for MSCA IF and ERC actions, we find significantly lower odds of success for applicants from EU-associated and third countries when their proposals are assessed by reviewers from third countries (0.40 and 0.25, respectively) (Table 7). Finally, for the SME instrument, we find that proposals submitted by applicants in non-private organisations have lower odds of success when assessed by reviewers affiliated with non-private organisations (Table 8).

These figures were developed using a multilevel logistic regression. Values provided indicate the 'adjusted odds ratio' (OR), which provides the likelihood of success given the applicant and reviewer characteristics. Values in Parentheses provide 95% confidence intervals – that is, there is a 95% likelihood that the value of the OR falls between the two values presented. The sample size for the analysis was 21 100 applications for the MSCA

IF group, 19 726 applications for the ERC group, and 65 059 applications for the SME group. We also provide p-values, which give an indication of the likely statistical significance of the difference between the odds of success for a particular group relative to the reference class. A lower p-value indicates greater significance. We indicate with stars the strength of this indication: *** p<0.001; ** p<0.01; * p<0.05.

Table 6: Outcome of mono-beneficiary grants by applicant and reviewer gender[12]

Action	Reviewer class	Applicant class	
		Female (reference class)	Male
MSCA	Female (reference class)	1	0.79 (0.71–0.87); p<0.001 ***
	Male	0.93 (0.84–1.04); p = 0.20	0.80 (0.72–0.89); p<0.001 ***
ERC	Female (reference class)	1	0.97 (0.84–1.12); p = 0.67
	Male	1.05 (0.89–1.24); p = 0.56	1.12 (0.96–1.31); p = 0.14
SME instrument	Female (reference class)	1	1.13 (1.01–1.27); p = 0.04 *
	Male	0.85 (0.70–1.03); p = 0.10	0.94 (0.79–1.11); p = 0.48

These figures were developed using a multilevel logistic regression. Values provided indicate the adjusted OR, which provides the likelihood of success given the applicant and reviewer characteristics. Values in brackets provide 95% confidence intervals – that is, there is a 95% likelihood that the value of the OR falls between the two values presented. The sample size for the analysis was 40 826 applications for the MSCA IF/ERC group and 65 059 applications for the SME group. We also provide p-values, which give an indication of the likely statistical significance of the difference between the odds of success for a particular group relative to the reference class. A lower p-value indicates greater significance. We indicate with stars the strength of this indication: *** p< 0.001; ** p< 0.01; * p<0.05.

[12] Multilevel logistic regression. OR adjusted odds ratio, 95% confidence intervals (LCI: lower confidence interval; UCI: upper confidence interval), p: p-value. MSCA IF n = 21 100. ERC n = 19 726. SME n = 65 059. *** p<0.001; ** p< 0.01; * p<0.05.

Table 7: Outcome of mono-beneficiary grants by applicant and reviewer nationality (for the purpose of this analysis, MSCA IF and ERC were combined, as they are both mono-beneficiary grants with similar objectives)				
Action	Reviewer class	Applicant class		
		EU Member State (reference class)	EU-associated country	Third country
MSCA IF and ERC	EU Member State (reference class)	1	0.67 (0.61–0.75); p<0.001 ***	0.72 (0.67–0.76); p<0.001 ***
	EU-associated country	0.84 (0.69–1.02); p = 0.08	0.96 (0.66–1.41); p = 0.85	0.85 (0.67–1.09); p = 0.19
	Third country	0.96 (0.87–1.07); p = 0.45	0.60 (0.49–0.73); p<0.001 ***	0.75 (0.66–0.84); p<0.001 ***
SME instrument	EU Member State (reference class)	1	1.19 (1.07–1.31); p = 0.001**	0.90 (0.80–1.00); p = 0.05
	EU-associated country	0.83 (0.61–1.14); p = 0.25	1.06 (0.66–1.72); p = 0.80	0.86 (0.52–1.42); p = 0.55
	Third country	1.09 (0.91–1.29); p = 0.34	1.11 (0.89–1.39); p = 0.36	0.89 (0.69–1.15); p = 0.38

These figures were developed using a multilevel logistic regression. Values provided indicate the adjusted OR, which provides the likelihood of success given the applicant and reviewer characteristics. Values in brackets provide 95% confidence intervals – that is, there is a 95% likelihood that the value of the OR falls between the two values presented. The sample size for the analysis was 40 826 applications for the MSCA IF/ERC group and 65 059 applications for the SME group. We also provide p-values, which give an indication of the likely statistical significance of the difference between the odds of success for a particular group relative to the reference class. A lower p-value indicates greater significance. We indicate with stars the strength of this indication: *** p< 0.001; ** p<0.01; * p<0.05.

Table 8: Outcome of mono-beneficiary grants by applicant and reviewer organisation (for the purpose of this analysis, MSCA IF and ERC were combined, as they are both mono-beneficiary grants with similar objectives).

Action	Reviewer class	Applicant class	
		Applicant in private organisation (reference class)	Applicant in non-private organisation
MSCA IF and ERC	Reviewer in private organisation (reference class)	1	1.18 (0.85–1.66); p = 0.32
	Reviewer in non-private organisation	0.56 (0.39–0.82); p = 0.003	0.88 (0.62–1.25); p = 0.47
SME instrument	Reviewer in private organisation (reference class)	1	0.50 (0.36–0.70); p<0.001 ***
	Reviewer in non-private organisation	0.98 (0.86–1.13); p = 0.83	0.543 (0.38–0.79); p = 0.001 **

From the literature we also identify other factors that are considered to affect the likelihood of success in application to EU Framework Programmes. In particular, Enger and Castellacci (2016) suggest that the two main factors increasing the likelihood of success are prior participation in EU Framework Programmes and scientific reputation (quality, impact and citation indices). This is reinforced by our bibliometric analysis of a sample of successful and unsuccessful applicants to ERC and MSCA. We find that successful applicants typically have a superior bibliometric track record (in terms of number of publications and citation indices) compared with unsuccessful applicants. This analysis also does not suggest bias against female applicants or those from less frequently participating countries. In fact, we find that successful female applicants, on average, have an inferior bibliometric track record compared with successful male applicants (Table 9). Similarly, successful applicants

from widening participation group countries[13], on average, have a weaker bibliometric track record than successful applicants from other Member States (Table 10).

Table 9: Bibliometric performance of successful male and female applicants

Pre-application indicators for successful applicants	Male	Female	Statistical significance (p-value)[14]
Average number of papers published per year	6.30	4.30	p<0.001
Average CNCI[15]	1.88	1.66	p<0.01
Average % of highly-cited papers[16]	24.0%	22.2%	p<0.01

Table 10: Summary of performance metrics pre-application by country group

Region	Number of successful applications in sample for bibliometric analysis	Pre-application average yearly output of papers	Pre-application average CNCI	Pre-application average % of highly-cited papers
EU Member States	1 804	5.6	1.79	22.8%
Widening participation group countries	83	4.0	1.39	18.5%
Other Member States	1 721	5.7	1.81	23.0%
EU-associated countries	136	6.8	1.76	23.7%

[13] Analysis of FP7 identified some Member States that had low participation rates, and these were targeted for additional support in H2020. These 'widening participation' countries are Bulgaria, Croatia, Cyprus, Czechia, Estonia, Hungary, Latvia, Lithuania, Luxembourg, Malta, Poland, Portugal, Romania, Slovakia and Slovenia.

[14] p-values give an indication of the likely statistical significance of the difference between the values for different groups. A lower p-value indicates greater significance and that the difference is meaningful.

[15] CNCI, or category-normalised citation impact, is the number of citations received by an applicant's papers, normalised to the average citation count for all papers published globally in the same year and field. The world average would be 1.00.

[16] Percent or proportion of an applicant's papers that are classified as highly-cited. In this report, a paper is considered highly-cited if it is among the top 10% most highly-cited papers published in the same year and field.

Looking at international practice, we find it notable that many funders (including the CFI, the DFG, the NCN, and the SNSF) undertake regular reviews of their processes to assess fairness in terms of differential outcomes for groups of applicants. Most seem to indicate a broadly equitable process, although specific examples of potential bias are highlighted by individual funders. For example, a 2017 analysis of SNSF found that male reviewers systematically rated male applicants higher than female applicants (Severin et al., 2020). Similarly, Innovate UK (part of UKRI) found that firm size had an impact on success rates, with smaller firms typically submitting more applications but having a lower success rate than larger firms, which was attributed to the enhanced administrative capabilities of large and medium-sized firms to develop applications (Salandra and Salter, 2019). In addition, analysis by SNSF of their processes found that peer reviewers selected by applicants consistently rated applications higher compared with reviewers identified by SNSF evaluation bodies (Severin et al., 2020). As a result, SNSF no longer allow applicants to nominate peer reviewers; they only applicants to provide exclusions. We also note that SNSF are exploring the use of double-blind review, in which the applicant is anonymised to the reviewer (see Section 3.6 for more details). Conducting periodic (e.g. annual) review of success rates by different groups and across different actions may be helpful for Horizon Europe to ensure that the process remains fair and to check for any emerging issues of concern.

We also see examples from different organisations of measures introduced that aim to reduce the risk of bias. One organisation, the CFI, have introduced bias training. At the moment, this is only required for CFI staff and is not compulsory for reviewers; however, the CFI are considering rolling this out as mandatory for reviewers in the future[17]. The DFG have taken different approaches to reduce bias, including the introduction of elected review boards, which are selected by the German scientific community (DFG, n.d.). These elected representatives are charged with quality-assuring the review process and ensuring reviewers have been appropriately selected, and this has received broad support from reviewers in a survey of their experiences. This corresponds to some elements of practice in H2020. For example, briefings have been provided to experts on unconscious and gender bias, and observers have been appointed to observe the proposal evaluation process, including such aspects as fairness and transparency.

2 Work is needed to encourage more female applicants to the framework programmes

Although the evaluation processes are broadly fair (as set out above) across H2020 applications, we find that the main imbalance in award holders is by gender, and that this is largely driven by the difference in application rates. Female applicants comprise only 28.7% of applicants to ERC and 17% of applicants to the SME instrument[18], although for MSCA IF, they comprise 40.8% of applicants (Table 11). Eurostat data suggests that women make up around 49.5% of the EU R&I workforce[19]. This discrepancy suggests that more work is needed to encourage and support female researchers and innovators to apply. Additionally, since the reasons for this imbalance are unclear, there may be value in conducting additional research to explore the reasons for lower rates of application among female researchers and innovators.

[17] From interview. See annex section 6.1.3 for more information.

[18] Note that this analysis is only possible for mono-beneficiary programmes, since data are held on the gender of the primary applicant only. ERC is also a mono-beneficiary programme, but gender data are not available for ERC applications.

[19] Eurostat, Employed HRST by sex, EU-28, average 2015–2019.

Table 11: Summary of applicant characteristics				
Variable	Levels	MSCA IF	ERC	SME instrument
n		64 519	57 185	65 059
Gender (%)	Female	26 333 (40.8%)	16 438 (28.7%)	11 079 (17.0%)
	Male	38 185 (59.2%)	40 747 (71.3%)	53 753 (82.6%)
	Gender not provided	1 (0.0%)	0 (0.0%)	227 (0.3%)
Nationality (%)	EU	33 780 (52.4%)	39 683 (69.4%)	48 886 (75.1%)
	EU-associated countries	4 509 (7.0%)	4 860 (8.5%)	8 251 (12.7%)
	Third countries	26 230 (40.7%)	12 642 (22.1%)	7 922 (12.2%)
Organisation (%)	Research organisations	5 437 (8.4%)	3 847 (6.7%)	10 140 (15.6%)
	Public bodies (excluding research organisations and secondary or higher education establishments)	28 343 (43.9%)	35 836 (62.7%)	38 746 (59.6%)
	Private for-profit entities (excluding higher and secondary education establishments)	12 701 (19.7%)	12 057 (21.1%)	0 (0.0%)
	Higher or secondary education establishments	251 (0.4%)	191 (0.3%)	407 (0.6%)
	Other	2 240 (3.5%)	854 (1.5%)	63 467 (97.6%)

The DFG have taken specific steps recently with the aim of improving female participation in their funding processes and reducing any gender bias. For example, the DFG recommend that all researchers, male and female, indicate in their CVs career gaps related to childcare, so that this can be considered by reviewers to potentially 'compensate for any disadvantages' (DFG, n.d.). Even if formal programme criteria are not met as a result of such delays, the DFG make clear that possible allowances can still be made (DFG, n.d.). For example, the Emmy Noether Programme, an early-career programme that enables post-doctorates with up to 4 years of experience to lead an independent junior research group, allows the years of experience to be extended by 2 years for every child under the

age of 12 (DFG, n.d.). Consideration of career breaks are already used in H2020 processes, for example in MSCA IF and ERC programmes.

The SNSF have also recently made changes to improve gender equality and representation in their evaluation and oversight groups (SNSF, 2021). According to new rules introduced in 2021, both men and women must account for at least 40% each of the Research Council and the Presiding Board. For commissions in specific fields where women are underrepresented, there is a requirement that female representation should be 20% higher than the proportion of professors who are female in that field. This is intended to ensure that female voices are represented in the decision-making processes. H2020 also included requirements for the proportion of female reviewers involved in calls.

We also note changes by the Canadian Institutes of Health Research (CIHR) in response to analysis of applications to a rapid response COVID-19 (coronavirus 19 disease) funding competition launched in February 2020, which received fewer female applicants and saw lower success rates among women than had been seen typically. Based on analysis of data, CIHR implemented a set of interventions for a subsequent call, which included increasing the application window; allowing the submission of abridged bio-sketches rather than the longer, online Canadian Common CV; creating a guidance document entitled *Why Sex and Gender Need to Be Considered in COVID-19 Research*; and requiring reviewers to assess the integration of demographic factors, such as gender, in research design (Witteman, Haverfield, and Tannenbaum, 2021). This highlights the need to monitor equality and bias in funding processes, since the specific context of the COVID-19 rapid response call created particular challenges for female applicants that had not been noted previously for other CIHR funding calls.

3 The majority of the burden of the proposal evaluation process falls on the applicants

Based on a survey of applicants to and reviewers for H2020, we have assessed the level of burden of the proposal preparation and evaluation processes. We find that applicants spend, on average, 27 days on a single-stage and 43 days on a multi-stage application (in cases where the application is subject to interview)[20]. Most of that time is spent on writing the applications – 85% of the time is spent on writing for single-stage applications, rising to 91% for multi-stage applications, reflecting the additional step. Reviewers spend around 4–6[21] days in total, collectively, reviewing each proposal. This increases to 5.5–7.5 days for processes that include an interview. This is overall time spent per application, taking into account the fact that applications will be reviewed by more than one reviewer at each stage of the process. Around half of that time is spent on the review of the proposal text and consensus meeting (20–35 hours, depending on whether it is a single- or multi-stage process), with panel meetings and interviews being less time consuming (each taking around 12 hours per application).

Comparing this to evidence from the literature (Table 12), we note that this appears to be somewhat longer than other funders' processes. However, comparator data are fairly limited and does not cover all the aspects of the process, in some cases focusing on a single aspect (e.g. peer review of written text only).

Overall, this means that around 80–85% of the burden falls on the applicants. This proportion is largely in line with the evidence from the literature, which suggests that for typical peer review–based proposal evaluation processes, around 75% of the burden falls

[20] Note that this is based on estimates by the lead applicant/contact point of their own time and their co-applicants' time spent on the application. Therefore, it may not be a completely accurate estimate, in particular in terms of time spent by co-applicants.

[21] Approximately 46 hours per application for multi-stage processes and 31 hours per application for single-stage processes, on average.

on applicants (Guthrie et al., 2018). This finding implies that efforts to reduce burden should focus first on applicants, rather than on reviewers, although of course all improvements to the efficiency of the process would be beneficial.

Despite this, we see relatively few examples of activities from our case studies of other funders that are directly intended to reduce burden on applicants. Several of the comparator organisations (including the CFI, the DFG, the SNSF and UKRI) use two-stage application processes, which are often intended to reduce burden. However, as set out below, we have reason to believe this may not be effective. Similarly, we see several examples where funders use page (or other text length) limits (including the CFI, the DFG, the NCN, the SNSF, UKRI and the National Research Foundation (NRF) of Singapore), which may in part aim to reduce burden on applicants (and reviewers). The limits vary significantly depending on funder and programme, but they seem to typically be in the range of 15–30 pages per application. However, the evidence from the literature suggests that this may not be effective in reducing time spent preparing applications (Barnett et al., 2015). This is reinforced by an evaluation conducted within UKRI of their Smart Grant programme, which found that burden was an issue for applicants, some of whom felt that the time input required for the application was disproportionate to the scale and nature of the award. In particular, applicants identified the strict character limits in application forms as a source of burden. In response to this concern, processes were changed so that applicants need no longer adhere to a character limit – however, a word limit for sections has been retained in its place. More widely in the literature, we see other examples of efforts to reduce burden on applicants. These include limits on the number of times an application can be submitted (Rockey, 2012), availability of multiple calls for applicants per year (Guthrie, Ghiga, and Wooding, 2018; Herbert et al., 2014), extending the application window (Bolli, 2014), and streamlining CV generation (Guthrie, Rodriguez-Rincon, et al., 2019a).

Table 12: Burden of different proposal evaluation processes			
Funding programme	Time spent by applicants[22]	Time spent by peer reviewers/panel members	Source
Applications submitted to Horizon 2020 funding schemes 2015–2020	25–47 days per application	5.5–7.5 days per application across all reviewers Initial review of proposal text takes c. 5–10 hours per reviewer	This study
Grant proposals submitted to the National Health and Medical Research Council (NHMRC) of Australia in 2009	22 days per application	-	Graves et al. (2011)
National Institutes of Health applications in 2015 from a school of nursing at a research-intensive academic health centre in the USA	9–20 days per application	-	Kulage et al. (2015)
Applications to Open Operating Grant Program of the NHMRC in 2009	21 days per application	Panel members are also peer reviewers c. 8 hours per application	Peckham et al. (2012)
Applications to four Australian Centre for Health Services Innovation health services research funding rounds, from May 2012 to November 2013	7 days per application	Peer reviewers are also panel members Less than 1 hour per application	Barnett et al. (2015)
Applications for NHMRC Project Grants in March 2012	34 days per application	-	Herbert et al. (2013)

4 Although the burden on reviewers is lower than that on applicants, it is worth considering the motivation and rewards for reviewers as a key part of the proposal evaluation process

Studies have also emphasised the burden of peer review on the reviewers in the context of grant applications (Bolli, 2014; Kitt, Williamson, and Paganetti, 2009; Snell, 2015; Guthrie, Ghiga, and Wooding, 2018). Reviewers are generally experts in their field who can critically assess the quality and feasibility of a research proposal. Therefore, being asked to be a reviewer in grant proposal evaluation can be considered an achievement, although reviewers generally do not receive academic recognition for their efforts in peer review (Schroter, Groves, and Højgaard, 2010). In addition, few institutions provide reviewers with protected time for peer review despite encouraging the process (Schroter, Groves, and Højgaard, 2010). It has been shown that reviewers are usually motivated to

[22] Time was converted to days, assuming 8 hours per day, where data were provided as hours.

participate in peer review by a sense of professional duty and fairness (Schroter, Groves, and Højgaard, 2010). However, the lack of recognition or (financial) reward for their time commitment has been shown to have a negative impact on reviewers' ability to deliver, with peer reviews often being conducted in the reviewer's own, personal time and participants pointing to their sense of professional duty and fairness as being the main reason that they continue to support peer review (Gluckman, n.d.; Schroter, Groves, and Højgaard, 2010).

In our analysis of the evaluators survey, we find that most respondents suggest that the fees they received do not reflect the time needed to evaluate the proposals. In terms of the time spent, our assessment is that the allowance made for the work is likely largely appropriate, although this may depend on how proposals are classified by the Commission (from simple to complex) and hence how much time is allowed. The average reviewer spends 7 hours per application for a single-stage process and 12 hours per application for a multi-stage process, assuming the application passes through both stages (including the proposal review and consensus meeting, but assuming no interview). Even for a simple proposal, the Commission model allows for around half a day to read any briefing material (for all applications reviewed) plus up to half a day for review and consensus meeting – and significantly more for more complex proposals, up to around 4.5 days for the most complex proposals (European Commission Directorate for Research and Innovation, n.d.). However, it may be that the hourly fee allowances are not appropriate, depending on the typical daily fee rates expected by individuals, which may differ depending on location and level of experience. The daily fee rate covered for H2020 reviewers is EUR 450, and the Commission indicates that this is intended to be an honorarium rather than fully covering the costs of individuals' time, although it is based on a per-hour model. In the future, there are plans to decouple the payment made from the time spent on the review to make it clearer that this is intended as an honorarium rather than a payment for time spent. This may help mitigate dissatisfaction with the level of compensation. We note in this context, as specified above, that although financial compensation may be important, reviewers are typically motivated by other, intrinsic factors. It may also be worth considering whether there could be other, non-financial ways of compensating reviewers for their contributions. One route could be through recognition of their role. For example, the use of Publons[23] to recognise the contribution of peer reviewers to journal publications is growing in popularity (Ortega, 2017). It may be that there is scope to adopt this type of recognition for grant peer review.

5 Two-stage application processes are typically intended to reduce burden, but they may actually increase overall burden

Our analysis finds that the time spent on preparing the full application is relatively similar for both single- and multi-stage proposals – excluding the first stage, the preparation of the full proposal takes an average of 23–28 days of applicant and co-applicant time (with the full proposal actually taking slightly longer for multi-stage proposals compared with single-stage) (Figure 2). This means that the 16 days spent preparing the first-stage application is additional effort and does not substitute any of the time that would be spent on the full application. Two-stage applications are typically used as a means to reduce burden, by reducing the number of applicants who have to complete a full application. However, considering the overall burden on applicants of the process, around 75–80% of applications would need to be rejected at the first stage for burden to be reduced – otherwise the first stage increases rather than reduces overall applicant burden. Based on our analysis, at present around 40% of applicants are rejected at the first stage. Assessing reviewer burden, we find that the same holds true: a significant proportion of applications

[23] Publons is a tool that allows tasks such as journal peer review and journal editing to be tracked and verified as part of a researcher's CV alongside more typical publication and citation metrics. https://publons.com/about/home/

(over 65%) would need to be rejected at the first stage for burden to be reduced – otherwise a two-stage assessment process increases rather than reduces reviewer burden.

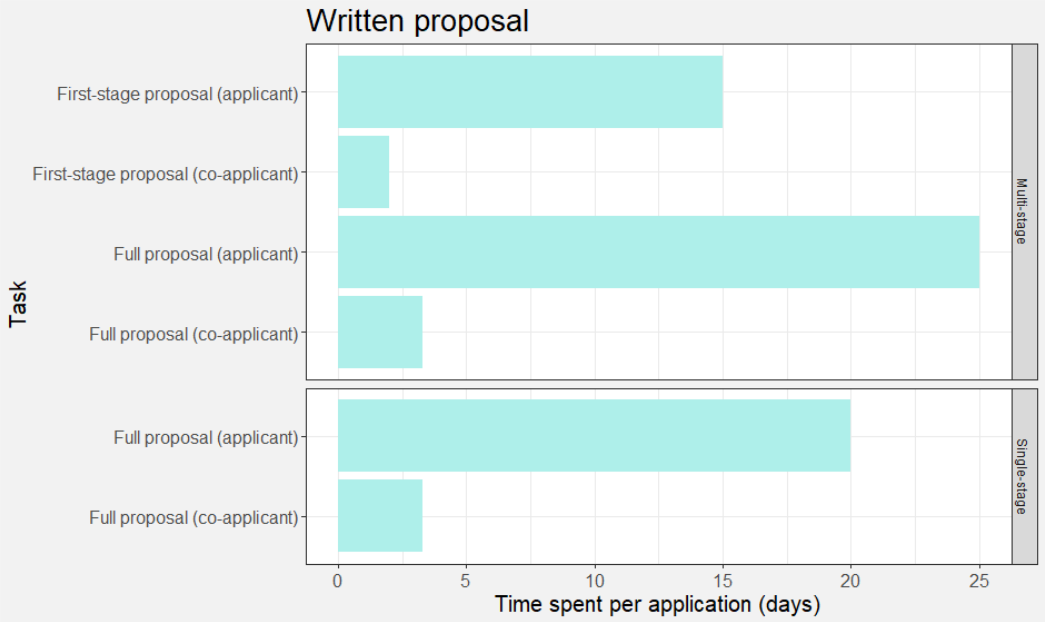


Figure 2: Median time spent on written proposal

Interestingly, several of the comparator organisations (including the CFI, the DFG, the SNSF and UKRI) use two-stage application processes, presumably in large part with the intention of reducing burden. However, we also note that the NRF (Singapore) use a two-stage process not just to address burden, but as a route to improve the quality of bids. Here, five-page 'white papers' from applicants receive feedback that can inform the preparation of their full proposals, which must also include a one-page rebuttal highlighting how the comments from reviewers have been addressed. This highlights that although two-stage processes may not be effective in reducing burden – and in fact may, overall, increase burden unless significant numbers of applications are eliminated at the first stage – there are other purposes a two-stage process can serve, and hence a two-stage process may still be appropriate in some cases.

6 Novel approaches, such as double-blind review and lotteries, could reduce bias and support more innovative research, but more evidence is needed

As well as adaptations to the existing proposal evaluation processes, there are more novel approaches that could be considered to support evaluations of funding. The SNSF (Switzerland) in particular have taken a more experimental approach to their funding, running pilots of two potential mechanisms to reduce bias: double-blinding and lotteries.

A double-blind review process (in which the applicant is anonymous to the reviewer as well as the reviewer being anonymous to the applicant) has been trialled by the SNSF for a new funding scheme, Spark, which aims to support 'unconventional ideas', taking a 'unique approach'. The process will be evaluated in 2021 but has been broadly considered successful, although subject to some teething troubles: technical challenges in achieving anonymisation resulted in 30% of applications to one call being insufficiently anonymised, and the call had to be re-run. More widely, double-blinding has also been trialled by the US National Science Foundation (NSF), in a process called 'The Big Pitch', which involved applicants submitting an anonymised two-page research proposal alongside a full conventional proposal (Bhattacharjee, 2012). Each application was assessed by a separate panel, and panels were asked to provide a list of their 'high-priority' proposals. The two lists differed greatly, and a comparative analysis showed that there was only 'a weak correlation' between the success outcomes of the two applications, as the panels'

assessments strongly diverged. This suggests that anonymisation can change assessment outcomes – and provides an early indication that the approach could be eliminating bias, although the evidence is not yet conclusive (Bhattacharjee, 2012). We also note that in calls for Pillar 2 under Horizon Europe, there will be a pilot of double-blind reviews for first-stage applications.

In addition to double-blind reviews, a lottery mechanism for funding has been piloted for one SNSF funding stream for postdoctoral mobility, building on several studies suggesting that lotteries could confer advantages in terms of reducing both bias and burden, and potentially in supporting more innovative research (Fang and Casadevall, 2016; Avin, 2019). The scheme situated a lottery process within a standard peer review process in which the best applications were selected for funding by expert reviewers and the least meritorious were rejected. The lottery aspect was used only for those applications 'in the middle', whose quality was not easy to distinguish. Ultimately, of the 134 applications reviewed, 8 (6%) proposals were decided by lottery. This means that although some fairness benefits may have been achieved, this approach has yet to realise some of the other potential benefits of a lottery approach, such as reduction in burden and timeline for assessment. However, this experiment does provide a useful proof of concept. Because the applicants were provided with sufficient information and rationale upfront, none of the affected applicants made any objection to the randomisation process. As a result of this experiment, SNSF are currently considering extending the use of a lottery approach to other funding schemes. A lottery approach has also been trialled by the New Zealand Science for Technological Innovation National Science Challenge (Adam, 2019). According to one of the members who oversee the Science for Technological Innovation fund, applicants are less disappointed by rejection if their application has been included in the ballot, as it means they were good enough to get funding but were unlucky at the draw (Adam, 2019). This finding is supported by wider survey evidence that researchers are supportive of lottery approaches, but more particularly where it is used to assess the 'middle' group of applications – with the outstanding applications already funded and the ineligible rejected, as in the SNSF approach (Liu et al., 2020).

7 More flexible use of consensus meetings may reduce reviewer burden and also enable more flexibility in the use of reviewers

Another option for reducing burden on reviewers is to limit consensus discussions to only those applications where there is disagreement among reviewers. The DFG (Germany) and the NHMRC (Australia) both take this approach, only discussing among reviewers in cases where the separate reviews have produced divergent results, thus reducing the time needed for those discussions – or, indeed, allowing the time available to be focused more on those applications where attention is needed. In 2019 and 2020 calls, several pilots of individual evaluation reports (IER) with no score were performed within MSCA, in which the score was only agreed during the consensus meeting after an agreement had been reached on the comments. Feedback received by the DG RTD stated that the consensus meetings were smoother and that there were fewer inconsistencies between scores and comments. The DG RTD is currently analysing whether the experts' feedback has evolved compared with previous evaluations using the IERs with scores.

A more radical approach could be to abandon consensus meetings altogether, and just use the average of scores received across reviewers. Evidence in the literature suggests than consensus and panel meetings may not make much difference to the consistency and reliability of funding decisions (Fogelholm et al., 2012; Pina, Hren, and Marušić, 2015; Carpenter et al., 2015; Herbert et al., 2015), although there are examples to the contrary (Baimpos et al., 2020). In our analysis of the burden of H2020 processes, we find that 4.6 hours of reviewer time is spent on preparing for and attending the consensus meeting for every application (taking into account the time of all reviewers involved) (Figure 3). This indicates that removing the consensus meeting process entirely could reduce reviewer burden by around 10–15%. In addition, in a survey of evaluators, 9% expressed dissatisfaction with existing consensus group processes.

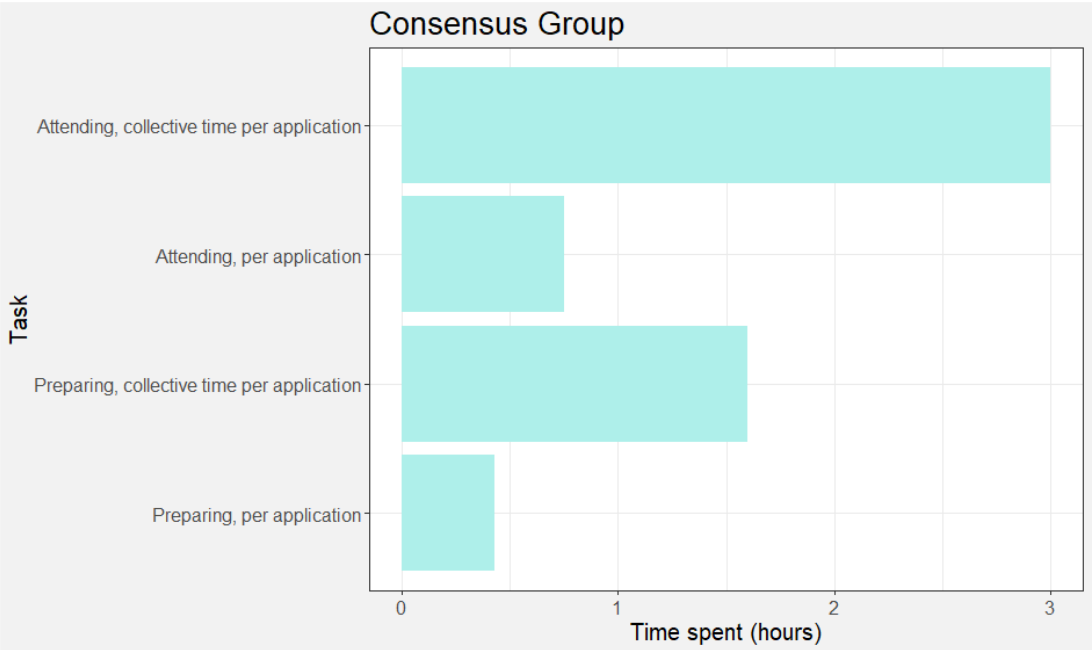


Figure 3: Median time spent on consensus group discussion

Another approach – which has largely been implemented in H2020 – is to move to remote (synchronous or asynchronous) consensus meetings. An analysis by Pina et al. (2021) of MSCA processes has shown that this has little impact on the final funding decision. Their conclusion is supported by wider evidence in the literature that suggests panel meetings can be conducted remotely with little impact on the outcomes of the process (Gallo, Carpenter, and Glisson, 2013). This approach has been adopted by several funders in the context of COVID-19, and both the NCN and the CFI have noted that this remote model, or a mixed model, is likely to be continued for panel meetings in the future, as they have not found the remote approach to affect the process significantly, and there are benefits both in terms of time and burden on reviewers and in terms of potential environmental benefits.

Using a remote and more targeted approach opens up opportunities to use reviewers in different ways. For example, it might be possible to use a greater number of reviewers. H2020 processes already use multiple reviewers to assess each proposal – typically four reviewers on average. Evidence from the literature is unclear on the optimal number of reviewers, with estimates ranging from 5 (Snell, 2015) to 11 (Graves, Barnett, and Clarke, 2011). Two examples from our case studies highlight funders that have chosen to increase the number of reviewers considering each proposal, to improve confidence in the process (the SNSF and the NHMRC). In the case of the NHMRC (Australia), this in part reflects the development and clarification of their evaluation framework for proposals. They have introduced a system in which different aspects of proposals may be evaluated by different people depending on their relevant expertise. For example, track record might be assessed by disciplinary experts (or potentially many of these, for a multidisciplinary team proposal), while such criteria as 'Synergy' and 'Knowledge gain' would be assessed by individuals with broad expertise. This type of approach could potentially be very helpful for assessing complex collaborative bids of the type commonly received through IA and RIA.

However, increasing the number of reviewers may not be without its own challenges. Several funders highlight the process of identifying and recruiting expert reviewers as a challenge. At the SNSF (Switzerland) this process is carried out by reviewers on their Research Council. To support this process, changes have been made, and now SNSF scientific officers provide an initial list of reviewers, to reduce the burden of identifying relevant individuals. Innovate UK, part of UKRI, have also looked at more novel ways of identifying reviewers, and in 2017, they issued a call for machine learning solutions to

improve their operations, including the matching of reviewers to proposals. Although the findings of the work have yet to be published or implemented, there is existing evidence of the potential to use text mining–based approaches to identify peer reviewers for journal articles (Arunachalam et al., 2013). Automated approaches to reviewer identification were also tried previously by the CIHR (Canada), as part of a wider package of changes to their funding system. Due to some challenges in implementation, these were not well received by the academic community (Gluckman et al., 2017). However, such approaches could still have value if implemented effectively and with support from the research community.

8 There may be scope to increase the clarity and consistency of application of funding criteria, for example through a clear framework for assessment and better alignment across actions

To ensure consistency in assessment of applications, it is important to have clarity and consensus on the rules and assessment criteria associated with the evaluation (Abdoul et al., 2012; Sattler et al., 2015; Gregorius et al., 2018). Evidence from the literature (focusing on ERC applications) suggests there may be a lack of clarity on how criteria – particularly the 'Excellence' criterion – should be implemented and that reviewers may not always be following action-specific guidance provided (Luukkonen, 2012; van den Besselaar, Sandström, and Schiffbaenker, 2018). This is because, although the criteria of 'Excellence', 'Impact', and 'Quality of implementation' are used across H2020, their specification varies by action, which may be a challenge where reviewers work across multiple actions.

These findings are supported by evidence from the reviewers survey, in which reviewers were asked to rate their level of satisfaction on different aspects of the evaluation process using a Likert-type scale[24]. Overall, levels of satisfaction with the process were very high, but consensus meetings and clarity on review criteria were both raised as areas of concern (although still only by a minority of respondents). In particular, lack of clarity on aspects of the rules – such as how the exceeding of page limits in the applications should be handled – were highlighted. Appropriateness, consistency of application, and weighting of evaluation of criteria were considered unsatisfactory by a small proportion of reviewers[25], and around 6% also had concerns about the scoring resolution process, expertise of reviewers and balance of expert backgrounds (Figure 4).

[24] Based on a survey of reviewers across all Horizon 2020 calls 2016/2017.

[25] These were considered unsatisfactory by the following proportion of respondents: Appropriateness – 4.2%, consistency of application – 6.5%, weighting of evaluation of criteria – 5.6%.

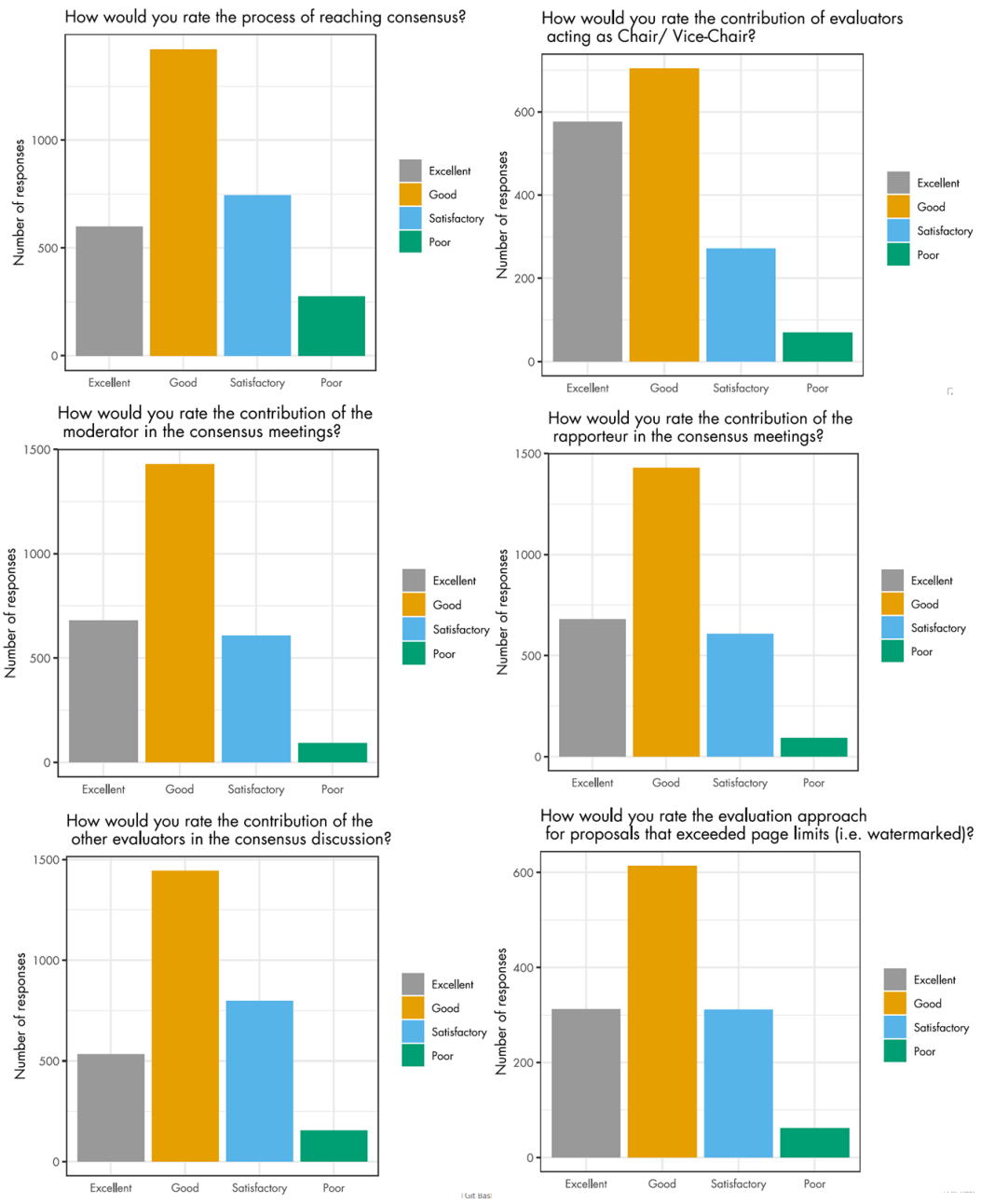


Figure 4: Survey answers on the level of satisfaction with consensus meetings

Several other funders provide examples of approaches that are intended to ensure more robust and consistent use of evaluation criteria in assessment. For example, applications to the CFI (Canada) are directly tailored to the criteria, with applicants being asked to directly state how their proposed work addresses those criteria as the bulk of the application. The NHMRC (Australia) have taken a different approach, developing a clear framework for the assessment of different criteria (e.g. applicant track record), and using this same framework for assessment across all of their funding schemes to promote consistency, clarity and reviewer alignment with the intended aims of the funder (see Figure 5). Innovate UK (part of UKRI) received feedback in an evaluation that the guidance to applicants from different reviewers could be inconsistent. To address issues with inconsistency among reviewers, Innovate UK introduced a moderation phase, in which outlying assessor scores are checked to improve overall consistency. The extent to which this has addressed the challenge of inconsistent feedback from reviewers is unclear, and it carries some potential downsides, since evidence from the literature suggests that high levels of reviewer disagreement may be indicative of innovative work (Guthrie, Rodriguez-Rincon, et al., 2019a; Guthrie, Rodriguez-Rincon, et al., 2019b).

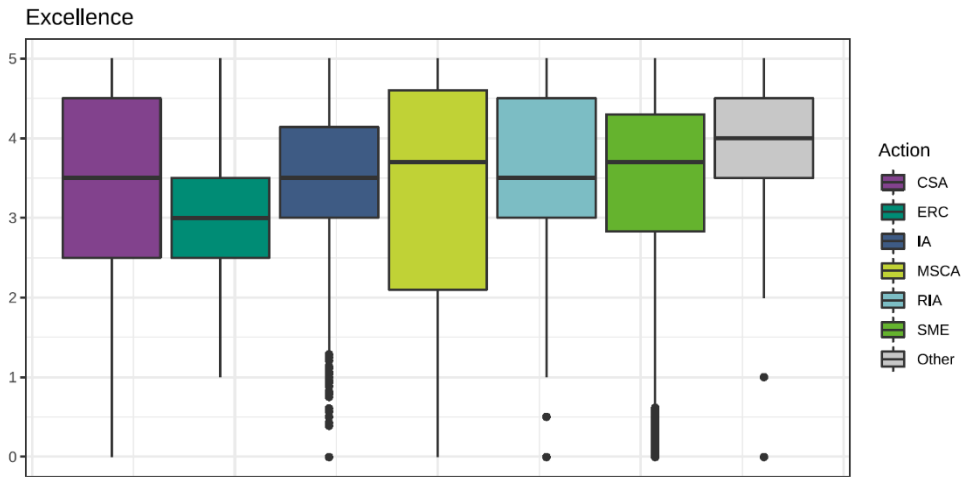
1. Publications (35%)
10 year list (taking Career Disruption into account)
Five best publications

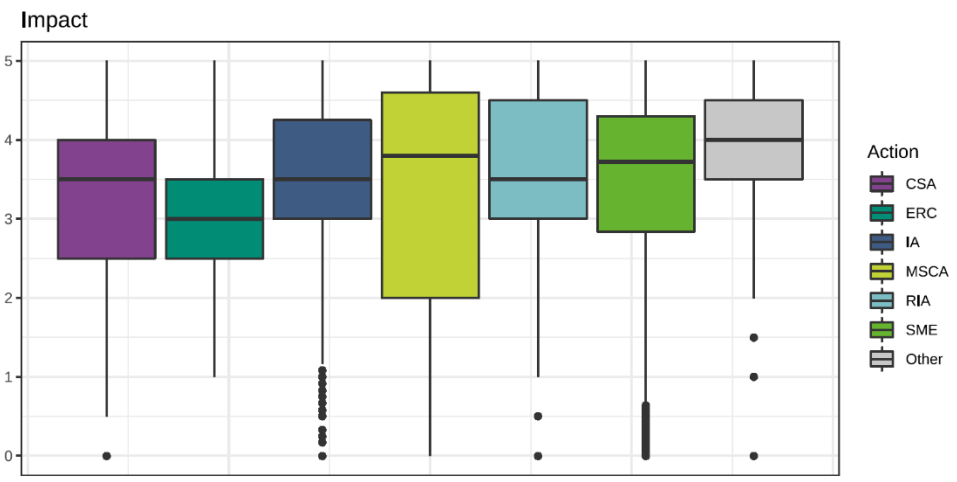
2. Research Impact (20%)			
Knowledge	Health	Economic	Social

3. Leadership (15%)
Research programs and team leadership
Institutional leadership
Research policy and professional leadership
Research mentoring

Figure 5: Framework for track record assessment used across NHMRC funding streams. Source: Kelso (2019)

We have analysed the scores of H2020 applications at the individual reviewer level to understand the extent to which we see consistency across reviewers and actions. Applications are assessed on a 0–5 scale, where 0 is 'not addressed', 1 is 'poor' and 5 is 'excellent', against three criteria: 'Excellence', 'Impact' and 'Quality and efficiency of implementation'. We find that median scores for 'Quality and efficiency of implementation' are higher, at 3.75, whereas the 'Excellence' and 'Impact' criteria have a median score of 3.5 (across all applications). Comparing actions, we find that ERC grants, overall, have the lowest median score across all criteria, while actions that make up the 'other' category – PCP (Pre-commercial Procurement), PPI (Public Procurement of Innovative Solutions), EJP (European Joint Programme) Cofund – have the highest median score across all criteria (Figure 6). This suggests that the scoring is not equivalent across criteria and actions, which suggests either difference in the quality of applications in relation to those criteria and actions or difference in the way in which the criteria are applied. As we noted above, the specification of criteria varies by action, which may be part of the reason for these variations.





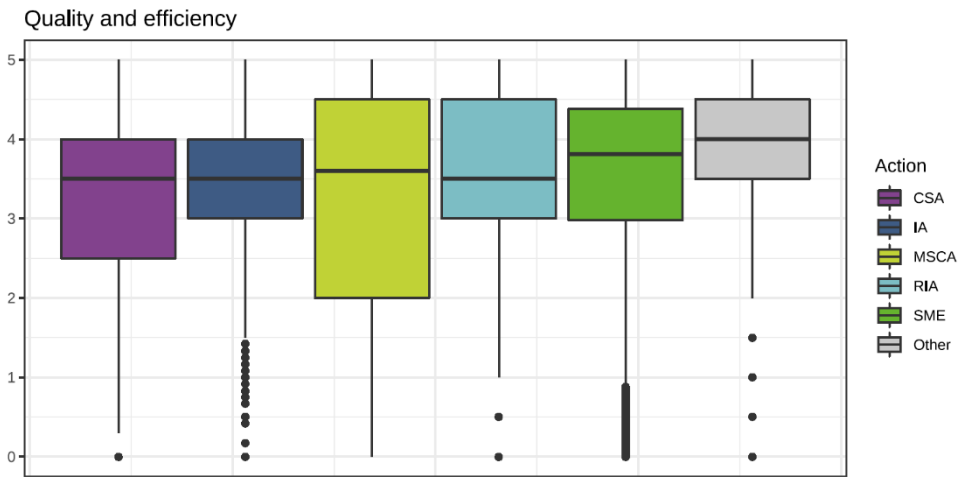


Figure 6: Median and range of scores by action and evaluation criterion

We also see some differences in scores based on the characteristics of individuals. The most 'generous' reviewers (those with a mean score across all reviews of 4 or higher) are predominantly male (60.8%), whereas the least generous reviewers (those with a mean score across all reviews of 1 or lower) have a more even gender balance (45.5% female and 54.5% male) and are predominantly under 50 years old (65.7%) (Table 13). This suggests there may be some differences and inconsistencies in how criteria are applied and scored between different evaluators.

Table 13: Characteristics of generous reviewers (scoring mean equal to or above 4) and less generous reviewers (scoring mean equal to or below 1)[26]			
Variable	Levels	Generous reviewers	Less generous reviewers
n		12 268	1 850
Gender (%)	Female	4 807 (39.2)	841 (45.5)
	Male	7 461 (60.8)	1 009 (54.5)
Age (%)	Under 50	5 778 (47.1)	1 215 (65.7)
	Over 50	6 490 (52.9)	635 (34.3)
Nationality (%)	EU	8 694 (70.9)	1 281 (69.2)
	EU-associated countries	751 (6.1)	123 (6.6)
	Third countries	2 823 (23.0)	446 (24.1)
Organisation (%)	Higher or secondary education establishment	4 921 (40.1)	843 (45.6)
	Private for-profit organisation	2 309 (18.8)	186 (10.1)
	Public organisation	936 (7.6)	291 (15.7)
	Research organisation	1 923 (15.7)	156 (8.4)
	Other	2 179 (17.8)	374 (20.2)

One approach used by several funders to ensure the quality and consistency of proposal evaluations is to recruit the pool of reviewers based on the nature of applications received, to establish a pool of reviewers with the necessary and appropriate experience to conduct the review adequately. This approach is used by the DFG (Germany) and the NCN (Poland), for example, and is also typical of H2020 processes. To expedite this recruitment, the CFI (Canada) ask institutions applying for their infrastructure funding to submit a 'notice of intent' with a brief description of the intended application (including suggested reviewers), to allow the CFI to recruit reviewers in advance of proposal submission. This step may help overcome some of the delays that can be associated with the reviewer identification process where this takes place after application submission, but it also places an additional step – and hence an additional burden – on applicants. It also avoids the need to ask applicants to suggest reviewers. As noted previously, reviewers suggested by applicants were found in an analysis by the SNSF to give higher scores.

[26] ERC Proof of Concept (PoC) grants were excluded from the scoring analysis because they are not scored based on a 1–5 scoring scale (as is the case for the Investigator (Starting, Consolidator and Advanced) and Synergy Grants), but, rather based on a pass/fail system.

9 ERC and MSCA funding processes effectively select candidates with better performance in terms of academic output both post- and pre-award, but proposal scores are a weak predictor of performance

Predictive power refers to the extent to which proposal evaluation processes effectively select the 'best' proposals, which go on to be the most successful. Defining success in a way that can be measured is a key challenge in assessing predictive power, and the main way this is usually done is through assessing bibliometric outcomes for successful compared with unsuccessful applications. This has limitations in that many aspects of the H2020 portfolio would not aim to achieve publications as a key outcome. Therefore, we limited analysis to ERC and MSCA IF actions. We found that a successful funding application is a good indicator of bibliometric performance in the years following award (Table 14). However, we also note that prior bibliometric performance is higher for successful applicants, so it may also be that this prior output is a good predictor both of success and of future publication numbers and quality. This hypothesis is in contrast to wider evidence from the literature, which is mixed, but which suggests that the predictive power of typical peer review processes is poor, with successful applications typically not showing significant differences in subsequent bibliometric performance compared with unsuccessful applications (van den Besselaar and Sandström, 2015; Danthi et al., 2015; Doyle et al., 2015; Fang, Bowen, and Casadevall, 2016; Kaltman et al., 2014; Reinhart, 2009). However, we also find that the overall score achieved by a proposal in the proposal evaluation process is only a weak predictor of performance (Figure 7). This suggests that, broadly speaking, ERC and MSCA IF processes are effective in picking the 'best' applicants (in terms of bibliometric output, and for the programmes mentioned) and eliminating the less meritorious, but, in line with most proposal evaluation processes, there is little to discern between the best unsuccessful proposals and those that are only just successful. This finding strengthens the case for consideration of alternative approaches, such as lotteries, for these borderline applications. We also note that the two streams in question are focused on academic excellence as a key selection criterion (indeed, the only selection criterion, typically, for ERC funding). This means that it is perhaps to be expected that outcomes of these processes are a better-than-average predictor of bibliometric performance when compared with other funding streams, which may target a wider range of outputs.

Table 14: Overall performance of successful and unsuccessful applicants post-award

	Average number of papers per year	Mean CNCI	Average % of highly-cited papers
Successful applicants	6.5	1.62	18.2%
Unsuccessful applicants	5.1	1.17	12.1%

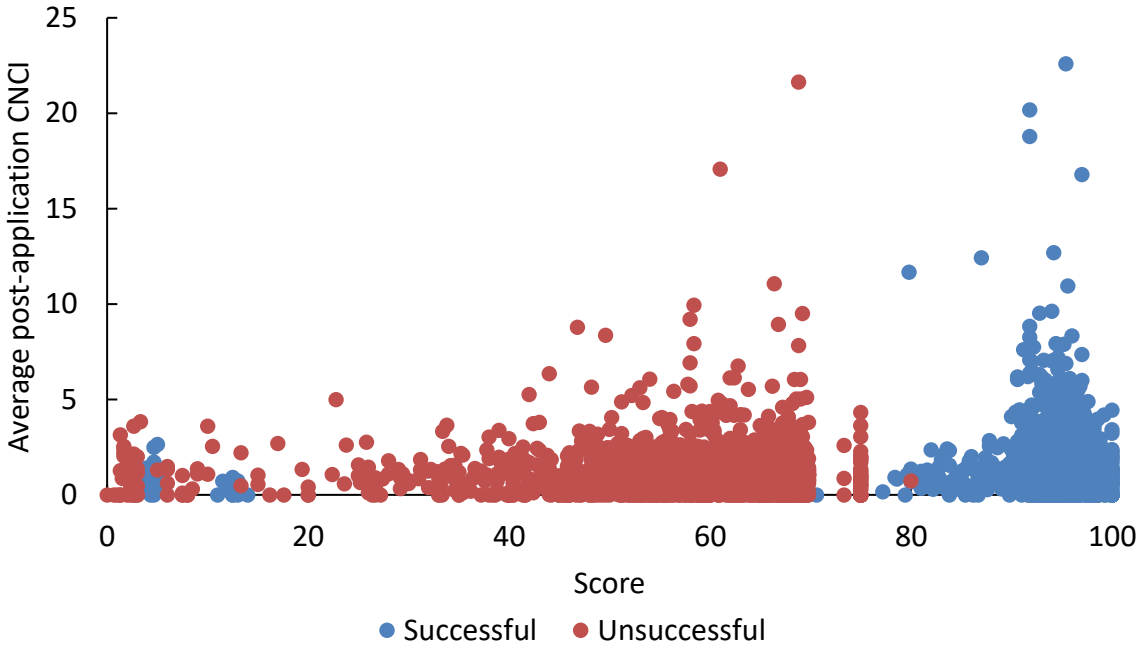


Figure 7: Relationship between applicants' category-normalised citation impact after application and the score given to their application

10 Horizon 2020 is broadly transparent compared with international benchmarks, but there may be scope to build on this transparency further, particularly at the pre-call stage

Transparency in proposal evaluation processes can be considered as the extent to which information associated with the process is disclosed (Ho et al., 2013; Murray et al., 2016; Tamblyn et al., 2018; Gurwitz, Milanesi, and Koenig, 2014). Transparency can have benefits not just in terms of openness and understanding of the process, but also in terms of impact on the perceived burden (Gurwitz, Milanesi, and Koenig, 2014). This is because it allows applicants to better understand the process that has led to the outcome of their application, and it allows reviewers to obtain recognition for their work. Assessing transparency against an established framework (Gurwitz, Milanesi, and Koenig, 2014), we find that the level of transparency is broadly similar across all H2020 actions and that, in general, it is greater than for many international comparators (Figure 8).

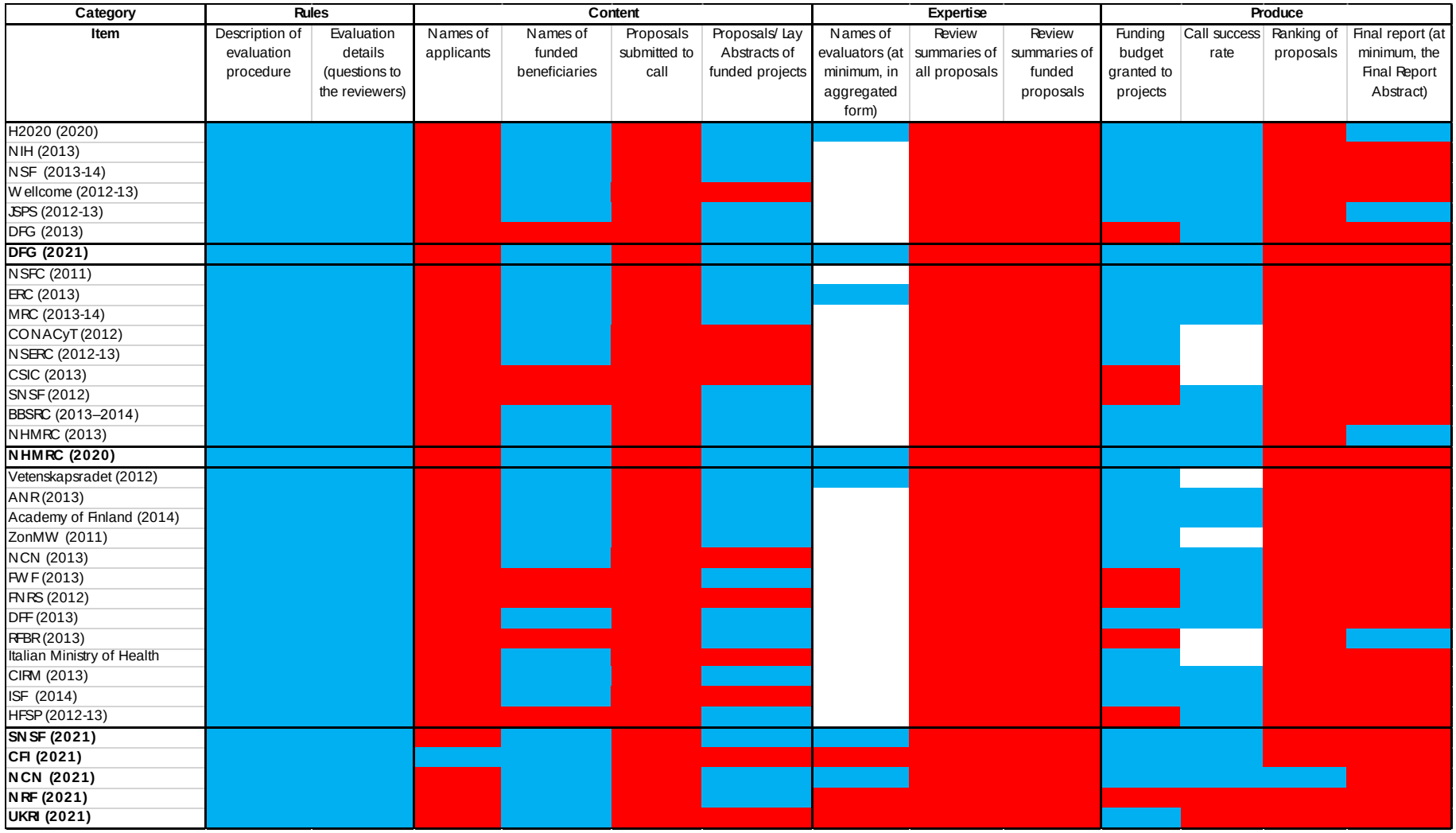


Figure 8: Transparency of H2020 compared with other funding agencies with regard to access to information; in blue, items that are publicly available, indicating an 'open', or transparent, system; in red, items that are not publicly available, indicating a 'closed' system; in white, items whose public availability is unknown

11 Peer review–based proposal evaluation processes tend to be conservative, and there are indications this may apply in the case of H2020

Conservatism and innovative R&I are difficult to define, which makes measuring the extent to which a proposal evaluation process supports innovative work difficult – and also makes designing a system to support innovative R&I challenging (Guthrie, Rodriguez-Rincon, et al., 2019a; Guthrie, Rodriguez-Rincon, et al., 2019b).

One approach to assessing conservatism proposed in the literature is to use reviewer disagreement as a proxy for innovation. This is an interesting approach – the idea is that reviewers are more likely to differ in opinion and understanding regarding more novel proposals and, indeed, may differ in their openness to these proposals depending on their risk appetite. Interestingly, a previous study of MSCA processes found that disagreement at the consensus meeting stage influenced the final score of a proposal negatively, resulting in lower scores (Pina, Hren, and Marušić, 2015). This may imply that more innovative proposals are disadvantaged by the need to reach consensus, which is built into the H2020 proposal evaluation process.

We have also tested this against the data we have available by looking at whether the number of reviewers and the level of disagreement among these reviewers impacts on the likelihood of proposal success. We find that higher levels of variation from mean scores (i.e. more disagreement) is associated with a lower likelihood of proposal success[27]. This merits further investigation but provides some initial suggestion that innovative (or at least controversial) proposals may be less likely to be successful and that the process may be conservative in the projects selected for funding.

One method proposed to support more innovative work is to include a formal assessment criterion on innovation or novelty in the proposal evaluation process (Liaw et al., 2017). This is the main approach we observe in our international case studies. One example is the NHMRC Ideas Grants programme, which includes a specific assessment of 'Innovation and creativity' as one of its assessment criteria and, unlike other NHMRC funding streams, retains panel meeting discussions, including the option for panel members to 'rescue' applications discarded in the first assessment round and bring them back into that meeting for discussion – allowing novel and controversial work to be considered by the panel rather than rejected early. The NRF also have specific funding streams focused on translational research, as well as assessment criteria dedicated to technical innovation – such as assessment of technology readiness levels. They also encourage applicants to consider impact metrics that go beyond academic outputs when developing their proposals. These approaches could be relevant to funds that seek to achieve translational or innovative results, such as the EIC Accelerator Pilot.

Taking a broader perspective, we note that it is also worth highlighting that, although the inconsistency in feedback to applicants is a potential area for improvement highlighted for Innovate UK, this inconsistency also reflects the independent nature of the review process: independent review occurs without a consensus phase, so at no point are assessor comments aggregated. Although inconsistency in feedback is in some ways a challenge, it might also support innovation, since consensus processes and discussion are typically conservative and may serve as a barrier to innovation.

We also see examples of approaches to support more innovative research in the literature. For example, the National Institutes of Health in the USA adopted an 'out-of-order funding'

[27] For example, ERC proposals with a medium and low coefficient of variation have 1.72 and 1.69 higher odds of being selected, respectively, than proposals with a high coefficient of variation. We see similar patterns for IA, RIA, CSA, MSCA and the SME instrument.

approach, in which a number of applications for innovative research are chosen for funding despite receiving lower scores than other funded research based purely on the peer review process (Lindner and Nakamura, 2015). The authors found that approximately 15% of funded applications were selected 'out of order', meaning that 15% of the funded applications included in the study would not have been funded if the decision had purely been based on peer review scores. In addition to the out-of-order funding approach, the NIH introduced various changes to the peer review process aimed at increasing the focus on innovation (Lindner et al., 2016), including reducing the length of the methodological proposal from 25 to 12 pages. Including innovation as a criterion for grant assessment could incentivise researchers to include innovative ideas and new approaches into their proposals (Guthrie, Ghiga, and Wooding, 2018). To this effect, the NIH asked applicants to detail what makes their application innovative. In addition, the NIH reduced the range of scoring options available to reviewers to five criteria ('Significance', 'Investigators', 'Innovation', 'Approach', and 'Environment') as well as having an overall impact score. The authors found that all the scored review criteria, including 'Innovation', were related to the overall impact score, and that good scores were necessary on all five scored review criteria to achieve a good overall impact score. Many funding agencies have also adopted the strategy of having a separate scheme to fund innovative research, allocating smaller funds with a shorter time frame to these specific streams, as suggested by Alberts (2009). For example, the Dutch organisation for health research and development, ZonMW, designed an 'off-road' programme seeking to fund high-risk, high-reward projects (Guthrie, Rodriguez-Rincon, et al., 2019a).

Overall, the evidence on the best ways to support more innovative research is inconclusive. The most commonly used approach is to use dedicated funding streams for innovative or risky research, but at present there is limited evidence on whether this approach is effective, and more evaluation is needed.

12 Feedback could be more focused on improvement and learning and more clearly structured around the evaluation criteria

Since most applications for any funding stream will be unsuccessful, feedback can be important in providing useful information to those unsuccessful applicants to help them learn and improve. In our survey, we asked applicants about the quality of feedback received. We found a mixed picture, with views on feedback varying between successful and unsuccessful applicants. Amongst successful applicants, 58% found the feedback received very or extremely useful, but among the unsuccessful applicants only 20% found it very or extremely useful (Figure 9). By contrast, 41% of unsuccessful applicants found the feedback not so useful or not at all useful, compared with 10% of successful applicants. This suggests that there may be scope for improvement, particularly in terms of communicating areas of improvement and learning to unsuccessful applicants.

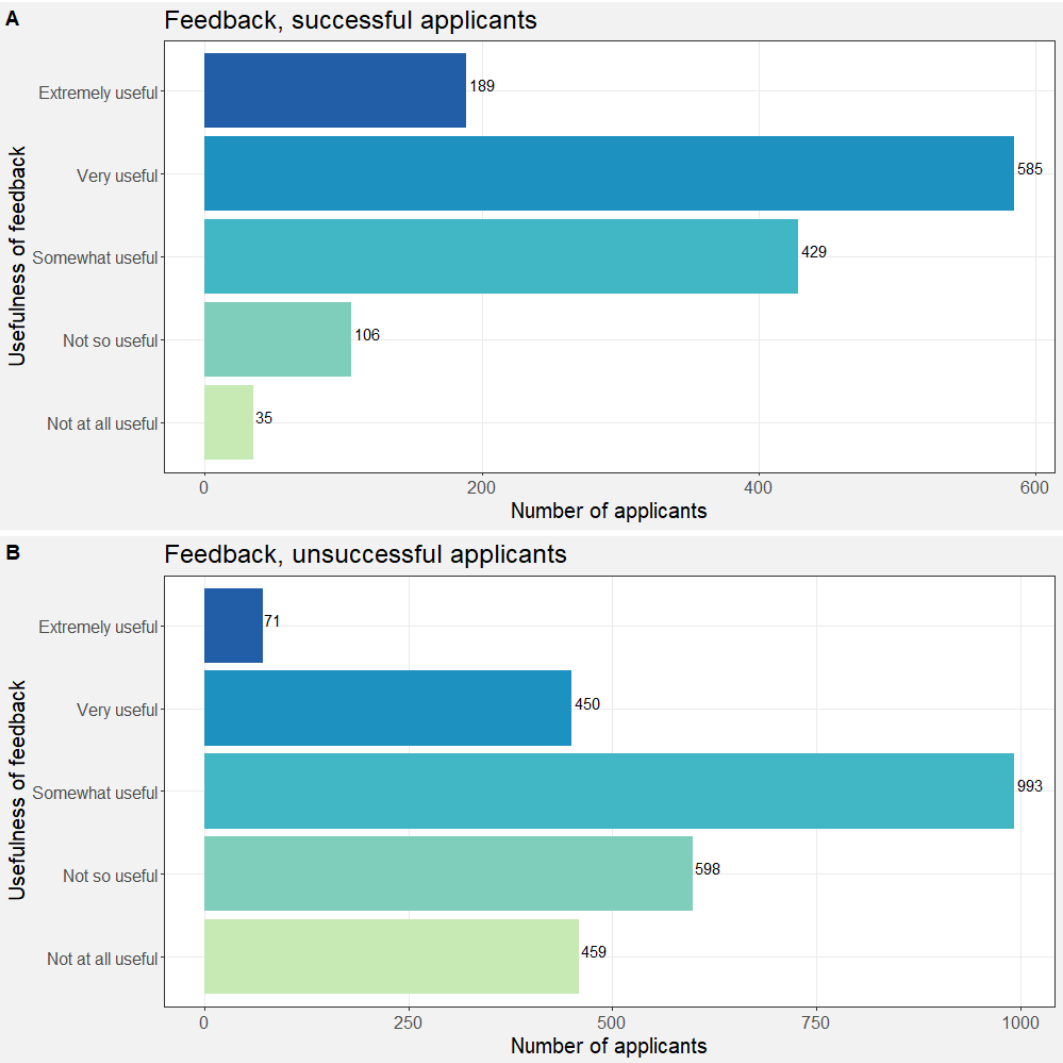


Figure 9: Applicant perception of the usefulness of feedback

Analysis of a sample of 300 feedback letters (termed evaluation summary reports) revealed several areas where there could be scope for improvements in the feedback provided to applicants. First, we find that for 40% of the ESRs, there were some inconsistencies in how the review was structured around the specified assessment criteria – typically with some criteria not fully addressed or with the review largely descriptive rather than analytical. We also find that 30.5% of the ESRs reviewed had some inconsistencies between the qualitative assessment provided and the scoring (Table 15). This is typically where the written review highlights specific weaknesses or is very positive but this is not reflected in the scoring. In terms of the constructiveness of the feedback, most reviews (90%) use constructive language, and almost all the rest used neutral language (Table 16). In addition, the vast majority (96%) are neutral in tone. One factor to note in this assessment is that reviewers are specifically advised not to provide guidance on how to improve the application; rather, they are advised just to assess its relative strengths and weaknesses. Therefore, it is positive that 90% of reviews were able to take a constructive tone.

Table 15: RAG[28] rating for alignment of scores and written assessment				
RAG rating for alignment between score and assessment	Number			Percentage
	Successful	Unsuccessful	Total	
Green	90	78	168	68.3%
Amber	32	43	75	30.5%
Red	2	1	3	1.2%
Total	124	122	246	100%

Table 16: Constructiveness of reviews				
Constructiveness	Number			Percentage
	Successful	Unsuccessful	Total	
Constructive	110	111	221	89.8%
Neutral	14	10	24	9.7%
Unconstructive	0	1	1	0.4%
Total	246		246	100%

Several of the case studies highlight examples where funders have actively sought to provide constructive feedback to inform future applications. For example, the CFI (Canada) provide detailed feedback on all assessment criteria (typically around 5–7 pages of feedback) after the first stage of the review process, to both successful and unsuccessful applicants. For those who progress, an additional 1–2 pages of feedback are provided after the second stage of the process. We note, however, that this does increase reviewer workload, and there are some programmes in which the feedback provided (as well as the initial application required from applicants) has been simplified, which may offer opportunities to streamline this process while still delivering good-quality feedback.

The NCN (Poland) also provide feedback to applicants rejected at the second stage, consisting of both the individual reviewer reviews and a consensus report compiling feedback. This is intended to offer both transparency on the decision made and constructive feedback to support applicants if they subsequently reapply. The NCN try to enable good-quality feedback by providing a structured assessment and evaluation form, which leads the reviewer through various components of feedback to be provided to the applicant, as

[28] RAG refers to red/amber/green.

well as online guidance and briefing sessions setting out expectations. This corresponds to H2020 processes, where there is an extensive briefing process for reviewers.

Innovate UK also provide applicants with feedback on their applications, which was generally positively received. However, there have been some criticisms regarding inconsistency of feedback among different reviewers. This is particularly challenging where changes have been made by applicants before resubmission and subsequent reviewers take a different perspective, creating the impression the requirements have shifted and generating uncertainty around the consistency of the process. To help address this, Innovate UK have tried to set expectations more clearly, making a clear statement that different assessors 'will have no prior knowledge of the original application or its feedback' and that addressing the initial assessors' feedback is no guarantee of funding. In addition, Innovate UK use a moderation step as noted above to try and improve consistency within a particular assessment process.

Some of the funders include processes where applicants not only receive feedback, but also are able to provide a rebuttal to that feedback, entering into a dialogue with reviewers. We noted above that this dialogue occurred between the first and second stages of application for NRF applications; the opportunity for rebuttal was introduced into SNSF processes several years ago, based on evaluation findings, in particular drawing on feedback from the Swiss academic community.

Overall, case studies illustrate that many funders provide feedback to applicants, but careful consideration of the content and format of this feedback is needed to ensure that the process is helpful and productive.

CHAPTER 4: RECOMMENDATIONS FOR HORIZON EUROPE

The key findings of our analysis of H2020 processes are summarised in the Executive Summary. Based on these observations, we identify a set of 10 recommendations for Horizon Europe.

1. **Conduct regular review of the fairness of the process:** Our analysis suggests that proposal evaluation processes are fair, but case studies suggest that it is good practice to review fairness regularly (e.g. on an annual basis) and present data openly to the R&I community. This would allow any issues emerging to be identified and addressed and would build confidence in the proposal evaluation process. One of the challenges here is the quality and availability of data to do this analysis thoroughly and across a range of characteristics. One option to improve data would be to collect fuller information on diversity and personal characteristics (e.g. including career stage, career breaks, mobility) on the application forms and to capture information on not just the lead applicants but also the co-applicants within larger consortium projects, so that analysis can be expanded beyond mono-beneficiary awards. Based on the analysis conducted by other funders, it seems to us that the collection of these types of information by other funding bodies is well established and that it allows them to monitor their proposal evaluation processes to ensure that they are promoting equality, diversity and inclusion and that they adapt where necessary.

2. **Encourage more female applicants:** Although proposal evaluation processes treat female applicants equitably, low application rates from female applicants relative to the wider R&I population drive low numbers of female award holders. Consideration should be given to ways to encourage greater numbers of female applicants, with examples from other funders and prior H2020 practice including the possibility to declare career breaks and have these considered within the assessment process, providing training and guidance to reviewers on consideration of gender, and ensuring appropriate participation of women in assessment panels and other leadership positions. However, there may also be wider systemic drivers of low application rates among female researchers and innovators. Further research to understand the reasons for lower levels of application among women, and how these issues could be addressed, would be valuable.

3. **Limit the use of multi-stage processes:** Multi-stage processes are typically intended to reduce burden by reducing the number of applicants who need to prepare a full application. However, our evidence suggests that unless 70% or more of applications are rejected, a multi-stage process increases burden on applicants overall. Therefore, multi-stage processes should only be used where a significant number of applicants are removed at the first stage, or where it confers other advantages to the process (e.g. through improving the quality of applications by providing feedback to applicants at an intermediate stage so they can make improvements). The likely demand could be judged based on prior experience in relation to a particular funding instrument, or through consultation and engagement with the sector. Based on our analysis, on average, around 40% of applications were rejected at the first stage in multi-stage processes for H2020, which suggests that in most cases a multi-stage process should not be used.

4. **Explore and experiment with novel approaches, such as double-blind review and lotteries:** Evidence from the literature and case studies suggests that novel approaches, such as lotteries, could have the potential to reduce bias and conservatism in proposal evaluation processes, and that these approaches are likely to be acceptable to the R&I community provided they are transparently implemented. Further, as proposal scores are only a weak predictor of bibliometric performance, this insight suggests there is a group of applications (at least within ERC and MSCA IF) where the use of lotteries may be appropriate. Until these innovative approaches are tested and explored in the real-world context, it will be difficult to know whether they can confer these advantages in practice; therefore experimentation – with robust evaluation – is

needed. These approaches may not be appropriate in all contexts and should be piloted on a smaller scale in the first instance. As Europe's flagship R&I funding mechanism, Horizon Europe should take the opportunity to offer leadership in the field and innovate in the proposal evaluation process.

5. **Streamline the use of consensus meetings:** Evidence suggests that the impact of consensus meetings on decision-making processes is limited and that it may be possible to reduce their use to only those cases where there is significant disagreement, particularly for less complex calls and funding mechanisms. Where they are used, a remote format (as currently used) is likely to save time and money and be more environmentally sustainable without any impact on outcomes. There may also be scope to conduct panel meetings remotely or reduce their role in a similar way.

6. **Improve the clarity of assessment criteria and ensure that the evaluation process is centred around them:** The understanding and implementation of evaluation criteria could be improved by providing clarity on how they should be understood and interpreted (perhaps including a clear and consistently used framework for assessment of the different criteria). The application form should also continue to be clearly structured around the criteria, as should the feedback provided to applicants, which could be better structured around the criteria.

7. **Consider more targeted use of reviewers depending on their expertise:** Improvements in alignment to assessment criteria and reduction in the use of consensus meeting may open up opportunities to use reviewers in a more targeted way to ensure that all aspects of applications can be assessed effectively. This could include increasing the number of reviewers in some cases and targeting their assessment to specific aspects of the proposal and criteria that align with their expertise. This could add particular value when assessing complex, multi-partner applications (e.g. through IA or RIA), allowing relevant experts to assess the expertise and excellence of aspects of the proposal associated with different disciplines or sectors. This may also make it easier to identify appropriate reviewers with relevant expertise. It may also be beneficial to select reviewers based on the specific topics of the proposal received rather than on broad disciplinary areas.

8. **Explore the scope to use novel technologies to improve the process:** Several new technologies could offer opportunities to enhance proposal evaluation processes. These include the use of artificial intelligence/machine learning approaches to identify appropriate reviewers, which can be a challenging and time-consuming process, and the use of recognition tools, such as Publons, to acknowledge and credit reviewers for their contribution to the process. Opportunities to pilot these technologies and explore the value they bring to support peer review–based processes, as well as any challenges in implementation, should be considered. One positive example here is the plan to use artificial intelligence to identify innovative proposals for the EIC accelerator in Horizon Europe. Rolling these new technologies out, however, should be done with care, since implementation problems can lead to a loss of trust within the community.

9. **Provide constructive feedback, especially for unsuccessful applicants:** Unsuccessful applicants are currently dissatisfied with the feedback they receive, and their level of satisfaction could be improved by providing more clear and constructive guidance on ways to improve where possible and by providing a clear and structured format to ensure that the feedback is directly aligned to the selection criteria. For example, a section on suggestions for improvement could be included in the evaluator's form, which would only be included in the feedback for unsuccessful applicants. This could ensure that for unsuccessful applicants, there is a useful outcome from the time-consuming application process. Filling out this section need not place an additional burden on reviewers if organised effectively, since this information is already required for the purposes of internal assessment.

10. **Consider routes to ensure innovative R&I is supported, such as targeted funding streams:** Peer review–based proposal evaluation processes are anti-

innovation, and there is evidence that this may apply in the case of Horizon 2020. The evidence on how to reduce conservatism in proposal evaluation processes is limited, but one commonly used approach is to create targeted funding streams specifically for innovative or risky R&I, often using specific assessment criteria focused on innovation or novelty to ensure assessment processes align with those aims. There may be scope to run specific streams of this nature within Horizon Europe, and if there is, their effectiveness in supporting 'riskier' R&I should be evaluated to improve knowledge on what works in funding innovative research.

REFERENCES

Abdoul, H., Perrey, C., Amiel, P., Tubach, F., Gottot, S., I. Durand-Zaleski, I. and Alberti, C. (2012), 'Peer review of grant applications: Criteria used and qualitative study of reviewer practices', *PLoS ONE*, Vol. 7, No 9, https://doi.org/10.1371/journal.pone.0046054.

Adam, D. (2019), 'Science funders gamble on grant lotteries', *Nature*, Vol. 575, No 7784, pp. 574-575, https://doi.org/10.1038/d41586-019-03572-7.

Adam, P., Ovseiko, P. V., Grant, J., Graham, K. E. A., Boukhris, O. F., Dowd, A., et al. (2018), 'ISRIA statement: Ten-point guidelines for an effective process of research impact assessment', *Health Research Policy and Systems*, Vol. 16, No 1, 8, https://doi.org/10.1186/s12961-018-0281-5.

Alberts, B. (2009), 'On incentives for innovation', *Science*, Vol. 326, No 5957, 1163, https://doi.org/10.1126/science.1184848.

Arunachalam, N., Sathya, E., Hismath Begum, S. and Uma Makeswari, M. (2013), 'An ontology based text mining framework for R&D project selection', *International Journal of Computer Science & Information Technology,* Vol. 5, No 1, https://doi.org/10.5121/IJCSIT.2013.5113.

Avin, S. (2019), 'Mavericks and lotteries', *Studies in History and Philosophy of Science Part A*, Vol. 76, August, pp. 13-23, https://doi.org/10.1016/j.shpsa.2018.11.006.

Baimpos, T., Dittel, N. and Borissov R. (2020), 'Unravelling the panel contribution upon peer review evaluation of numerous, unstructured and highly interdisciplinary research proposals', *Research Evaluation,* Vol. 29, No 3, pp. 316-326, https://doi.org/10.1093/reseval/rvz013.

Barnett, A. G., Herbert, D. L., Campbell, M., Daly, N., Roberts, J. A., Mudge, A. and Graves, N. (2015), 'Streamlined research funding using short proposals and accelerated peer review: An observational study', *BMC Health Services Research,* Vol. 15, No 1, 55, https://doi.org/10.1186/s12913-015-0721-7.

Besselaar, P. van den and Sandström, U. (2015), 'Early career grants, performance, and careers: A study on predictive validity of grant decisions', *Journal of Informetrics,* Vol. 9, No 4, pp. 826-838, https://doi.org/10.1016/j.joi.2015.07.011.

Besselaar, P. van den, Sandström, U. and Schiffbaenker, H. (2018), 'Studying grant decision-making: A linguistic analysis of review reports', *Scientometrics,* Vol. 117, No 1, pp. 313-329, https://doi.org/10.1007/s11192-018-2848-x.

Bhattacharjee, Y. (2012), 'NSF's "Big Pitch" tests anonymized grant reviews', *Science,* Vol. 336, No 6084, pp. 969-970, https://doi.org/10.1126/science.336.6084.969.

Bolli, R. (2014), 'Actions speak much louder than words', *Circulation Research,* Vol. 115, No 12, pp. 962-966, https://doi.org/10.1161/CIRCRESAHA.114.305556.

Carpenter, A. S., Sullivan, J. H., Deshmukh, A., Glisson, S. R. and Gallo, S. A. (2015) 'A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels', *BMJ Open,* Vol. 5, No 9, p. e009138, https://doi.org/10.1136/bmjopen-2015-009138.

Danthi, N. S., Wu, C. O., DiMichele, D., Hoots, W. K. and Lauer, M. S. (2015), 'Citation impact of NHLBI R01 grants funded through the American Recovery and Reinvestment Act as compared to R01 grants funded through a standard payline', *Circulation Research,* Vol. 116, No 5, pp. 784-788, https://doi.org/10.1161/CIRCRESAHA.116.305894.

DFG (n.d.a) 'DFG, German Research Foundation – Assessing individual circumstances', accessed 28 January 2021, https://www.dfg.de/en/research_funding/principles_dfg_funding/diversity/people_with_disabilities/individual_circumstances/index.html.

———. (n.d.b), 'DFG, German Research Foundation – FAQ: Emmy Noether Programme', accessed 28 January 2021, https://www.dfg.de/en/research_funding/faq/faq_emmy_noether/index.html.

———. (n.d.c), 'DFG, German Research Foundation – General information on dealing with diversity in the processing of proposals', accessed 28 January 2021, https://www.dfg.de/en/research_funding/principles_dfg_funding/diversity/diversity_allg/index.html.

———. (n.d.d), 'DFG, German Research Foundation – Membership period 2020–2024', accessed 28 January 2021, https://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/membership_2020_2024/index.html.

Doyle, J. M., Quinn, K., Bodenstein, Y. A., Wu, C. O., Danthi, N. and Lauer, M. S. (2015), 'Association of percentile ranking with citation impact and productivity in a large cohort of de novo NIMH-Funded R01 grants', *Molecular Psychiatry,* Vol. 20, No 9, pp. 1030-1036, https://doi.org/10.1038/mp.2015.71.

Enger, S. G. and Castellacci, F. (2016), 'Who gets Horizon 2020 research grants? Propensity to Apply and Probability to Succeed in a Two-Step Analysis', *Scientometrics,* Vol. 109, No 3, pp. 1611-1638.

Fang, F., Bowen, A. and Casadevall, A. (2016), 'NIH peer review percentile scores are poorly predictive of grant productivity', *ELife,* Vol. 5, February, p. e13323, https://doi.org/10.7554/eLife.13323.

Fang, F. and Casadevall, A. (2016), 'Research funding: The case for a modified lottery', *MBio,* Vol. 7, No 2, https://doi.org/10.1128/mBio.00422-16.

Fogelholm, M., Leppinen, S., Auvinen, A., Raitanen, J., Nuutinen, A. and Väänänen, K. (2012), 'Panel discussion does not improve reliability of peer review for medical research grant proposals', *Journal of Clinical Epidemiology,* Vol. 65, No 1, pp. 47-52, https://doi.org/10.1016/j.jclinepi.2011.05.001.

Gallo, S. A., Carpenter, A. S. and Glisson, S. R. (2013), 'Teleconference versus face-to-face scientific peer review of grant application: Effects on review outcomes', *PLoS ONE*, Vol. 8, No 8, p. e71693, https://doi.org/10.1371/journal.pone.0071693.

Gluckman, P. (2012), 'Which science to fund: Time to review peer review?', Office of the Prime Minister's Science Advisory Committee, London.

Gluckman, P., Ferguson, M., Glover, A., Grant, J., Groves, T., Lauer, M. S. and Ulfendahl, M. (2017), 'International Peer Review Expert Panel report: A report to the Governing Council of the Canadian Institutes of Health Research – CIHR', Canadian Institutes of Health Research, Ottawa, https://cihr-irsc.gc.ca/e/50248.html.

Grant, J. (2021), 'Academic incentives and research impact: Developing reward and recognition systems to better people's lives', AcademyHealth. https://academyhealth.org/publications/2021-02/changing-academic-research-social-good

Graves, N., Barnett, A., and Clarke, P. (2011) Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. BMJ. Vol. 343. 10.1136/bmj.d4797

Gregorius, S., Dean, L., Cole, D. C. and Bates, I. (2018), 'The peer review process for awarding funds to international science research consortia: A qualitative developmental evaluation', *F1000Research,* Vol. 6, January, pp. 1808, https://doi.org/10.12688/f1000research.12496.3.

Gurwitz, D., Milanesi, E. and Koenig, T. (2014), 'Grant application review: The case of transparency', *PLoS Biology,* Vol. 12, No 12, https://doi.org/10.1371/journal.pbio.1002010.

Guthrie, S., Ghiga, I. and Wooding, S. (2018), 'What do we know about grant peer review in the health sciences?' *F1000Research,* Vol. 6, March, https://doi.org/10.12688/f1000research.11917.2.

Guthrie, S., Rodriguez-Rincon, D., McInroy, G., Ioppolo, B. and Gunashekar, S. (2019a), 'Measuring bias, burden and conservatism in research funding processes', *F1000Research,* Vol. 8, June, 851, https://doi.org/10.12688/f1000research.19156.1.

Guthrie, S., Rodriguez-Rincon, D., McInroy, G., Ioppolo, B. and Gunashekar, S. (2019b), Design and development of an evaluation framework for the NHMRC: Review of International practice, overview of NHMRC's data and grant management systems, and an evaluation framework and metrics*,* RAND Corporation, Santa Monica, CA, https://www.rand.org/pubs/research_reports/RR2984.html.

Herbert, D. L., Barnett, A. G., Clarke, P. and Graves, N. (2013). On the time spent preparing grant proposals: An observational study of Australian researchers. BMJ Open, 3(5). doi:10.1136/bmjopen-2013-002800

Herbert, D. L., Coveney, J., Clarke, P., Graves, N. and Barnett, A. G. (2014), 'The impact of funding deadlines on personal workloads, stress and family relationships: A qualitative study of Australian researchers', *BMJ Open,* Vol. 4, No 3, p. e004462, https://doi.org/10.1136/bmjopen-2013-004462.

Herbert D. L., Graves, N., Clarke, P., et al. (2015), Using simplified peer review processes to fund research: A prospective study, *BMJ Open,* Vol. 5, p. e008380, https://doi.org/10.1136/bmjopen-2015-008380.

Ho, R. Chun-Man, Mak, K.-K., Tao, R., Lu, Y., Day, J. R. and Pan, F. (2013), 'Views on the peer review system of biomedical journals: An online survey of academics from high-ranking universities', *BMC Medical Research Methodology*, Vol. 13, No 1, 74, https://doi.org/10.1186/1471-2288-13-74.

Kaltman, J., Evans, F., Danthi, N., Wu, C., DiMichele, D. and Lauer M. (2014), 'Prior publication productivity, grant percentile ranking, and topic-normalized citation impact of NHLBI Cardiovascular R01 grants', *Circulation Research*, Vol. 115, No 7, pp. 617-624, https://doi.org/10.1161/CIRCRESAHA.115.304766.

Kelso, A. (2019), 'Peer review for NHMRC's new grant program', https://www.tri.edu.au/event/nhmrc-funding-video.

Kitt, C., Williamson, J. and Paganetti, H. (2009), 'TH-A-BRD-01: The future of NIH research funding', *Medical Physics*, Vol. 36, No 6, Part 26, pp. 2791-2791, https://doi.org/10.1118/1.3182589.

Kulage, K. M., Schnall, R., Hickey, K. T., Travers, J., Zezulinski, K.,Torres, F. et al. (2015), 'Time and costs of preparing and submitting an NIH grant application at a school of nursing'. *Nursing Outlook*, Vol. 63, No 6, pp. 639-649. https://doi.org/10.1016/j.outlook.2015.09.003.

Liaw, L., Freedman, J. E., Becker, L. B., Mehta, N. N. and Liscum, L. (2017), 'Peer review practices for evaluating biomedical research grants: A scientific statement from the

American Heart Association', *Circulation Research*, Vol. 121, No 4, pp. e9-19, https://doi.org/10.1161/RES.0000000000000158.

Lindner, M. D. and Nakamura, R. K. (2015), 'Examining the Predictive Validity of NIH Peer Review Scores', *PLoS ONE*, Vol. 10, No 6, p. e0126938, https://doi.org/10.1371/journal.pone.0126938.

Lindner, M. D., Vancea, A., Chen, M.-C. and Chacko, G. (2016), 'NIH peer review: Scored review criteria and overall impact', *American Journal of Evaluation*, Vol. 37, No 2, pp. 238-249, https://doi.org/10.1177/1098214015582049.

Liu, M., Choy, V., Clarke, P., Barnett, A., Blakely, T. and Pomeroy, L. (2020), 'The acceptability of using a lottery to allocate research funding: A survey of applicants', *Research Integrity and Peer Review*, Vol. 5, No 1, 3, https://doi.org/10.1186/s41073-019-0089-z.

Luukkonen, T. (2012), 'Conservatism and risk-taking in peer review: Emerging ERC practices', *Research Evaluation*, Vol. 21, No 1, pp. 48-60, https://doi.org/10.1093/reseval/rvs001.

Marques, M., Powell, J. J. W., Zapp, M. and Biesta, G. (2017), 'How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014', *European Educational Research Journal*, Vol. 16, No 6, pp. 820-842, https://doi.org/10.1177/1474904117730159.

Murray, D. L., Morris, D., Lavoie, C., Leavitt, P. R., MacIsaac, H., Masson, M. E. J. and Villard, M.-A. (2016), 'Bias in research grant evaluation has dire consequences for small Universities', *PLoS ONE*, Vol. 11, No 6, p. e0155876, https://doi.org/10.1371/journal.pone.0155876.

Ortega, J. L. (2017), 'Are peer-review activities related to reviewer bibliometric performance? A Scientometric Analysis of Publons', *Scientometrics*, Vol. 112, No 2, pp. 947-962.

Peckham D., Bosompra, K. N. and Manuel, C. (2012). Evaluation of the open operating grant program, final report 2012. Ottawa: http://www.cihr-irsc.gc.ca/e/45846.htmlPina, D. G., Hren, D., and Marušić, A. (2015), 'Peer review evaluation process of Marie Curie Actions under EU's Seventh Framework Programme for Research', *PLoS ONE*, Vol. 10, No 6, p. e0130753, https://doi.org/10.1371/journal.pone.0130753.

Reinhart, M. (2009), 'Peer Review of grant applications in biology and medicine: Reliability, fairness, and validity', *Scientometrics*, Vol. 81, No 3, pp. 789-809, https://doi.org/10.1007/s11192-008-2220-7.

Rockey, S. (2012), 'The A2 resubmission policy continues: A closer look at recent data', NIH Extramural Nexus (blog). 28 November 2012, https://nexus.od.nih.gov/all/2012/11/28/the-a2-resubmission-policy-continues-a-closer-look-at-recent-data/.

Salandra, R. and Salter, A. (2019), 'Helping Innovate UK understand their grant assessment process', Innovation Caucus (blog). 17 December 2019, https://innovationcaucus.co.uk/2019/12/17/helping-innovate-uk-understand-their-grant-assessment-process/.

Sattler, D. N., McKnight, P. E., Naney, L. and Mathis, R. (2015), 'Grant peer review: Improving inter-rater reliability with training', *PLoS ONE*, Vol. 10, No 6, p. e0130450, https://doi.org/10.1371/journal.pone.0130450.

Schroter, S., Groves, T. and Højgaard, L. (2010), 'Surveys of current status in biomedical science grant review: Funding organisations' and grant reviewers' perspectives', *BMC Medicine*, Vol. 8, October, 62, https://doi.org/10.1186/1741-7015-8-62.

Severin, A., Martins, J., Heyard, R., Delavy, F., Jorstad, A. and Egger, M. (2020), 'Gender and other potential biases in peer review: Cross-Sectional Analysis of 38 250 external peer review reports', *BMJ Open*, Vol. 10, No 8, p. e035058, https://doi.org/10.1136/bmjopen-2019-035058.

Snell, R. R. (2015), 'Menage a quoi? Optimal number of peer reviewers', *PLoS ONE*, Vol. 10, No 4, https://doi.org/10.1371/journal.pone.0120838.

SNSF (2021), 'Quotas to Promote gender equality in research – SNF', 5 February 2021, http://www.snf.ch/en/researchinFocus/newsroom/Pages/news-210205-quotas-to-promote-gender-equality-in-research.aspx.

Tamblyn, R., Girard, N., Qian, C. J. and Hanley, J. (2018), 'Assessment of potential bias in research grant peer review in Canada', *Canadian Medical Association Journal*, Vol. 190, No 16, pp. 489−499, https://doi.org/10.1503/cmaj.170901.

Witteman, H. O., Haverfield, J. and Tannenbaum, C. (2021), 'COVID-19 gender policy changes support female scientists and improve research quality', *Proceedings of the National Academy of Sciences*, Vol. 118, No 6, https://doi.org/10.1073/pnas.2023476118.

Wooding, S. and Grant, J. (2003), Assessing research: The researchers' view, RAND Corporation, Santa Monica, CA, https://www.rand.org/pubs/monograph_reports/MR1698.html.

Getting in touch with the EU

IN PERSON
All over the European Union there are hundreds of Europe Direct information centres.
You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

ON THE PHONE OR BY EMAIL
Europe Direct is a service that answers your questions about the European Union.
You can contact this service:
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696, or
– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

ONLINE
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU PUBLICATIONS
You can download or order free and priced EU publications from: https://op.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU LAW AND RELATED DOCUMENTS
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU
The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

The interim evaluation of Horizon 2020, the report by the Lamy Group, and other stakeholder feedback, noted that the current system for evaluating proposals is in general appreciated and one of the reasons for widely acknowledged successful implementation of Horizon 2020. With a view of maintaining and further developing a modern, world-class system, and identifying further simplifications where possible, the study assessed the performance of the Horizon 2020 proposal submission and evaluation process, including by taking into account international practices.

Studies and reports

Publications Office
of the European Union