Australian Government
National Health and Medical Research Council

Research Impact Track Record Assessment (RITRA) framework

Evaluation Report

30 January 2024

BUILDING
A HEALTHY
AUSTRALIA

Contents

Executive summary

Background

NHMRC's Research Impact Track Record Assessment (RITRA) framework requires researchers to report on past research impacts in their applications for Investigator and Synergy Grants, and requires peer reviewers to assess and score these reported impacts.

Implementation of this framework is intended to provide an incentive for researchers to consider future impact when planning and conducting research, ideally leading to an increase in the translation of NHMRC-funded research and improved public health.

This report describes the results of a process evaluation that sought to determine whether the RITRA framework has been implemented as intended.

Questions and findings

The evaluation questions and associated findings are as follows:

A1 – How easy was it to provide the impact text?

Applicants found it difficult to provide the impact text and both applicants and peer reviewers thought that there was substantial overlap in the text provided in the three impact sub-sections. Applicants found the examples of evidence to support their impact statements provided by NHMRC helpful but thought that they could be improved.

A2 – Are the impact types useful?

Most applicants consider that the four impact types, knowledge, health, social and economic, allow them to report all the impacts that they would like to, however some peer reviewers raised concerns that applicants had selected incorrect impact types.

Data analysis confirms that, across all applications, the impact types selected do relate to the impact text being provided, however applicants report knowledge impacts even if they have not selected this impact type.

A3 – What types of impacts are being reported?

Applicants more often report on knowledge impact than they do on benefits experienced by stakeholders beyond the research sector.

A4 – Is the impact text duplicating the publications text?

The impact and publications texts provided by applicants overlap substantially.

B1 – How easy was it to assess the impact text?

Peer reviewers expressed mixed views about how easy the impact text was to assess but were more united in finding the Category Descriptors to be unhelpful and offered a variety of suggestions for their improvement.

B2 – Do applicant characteristics influence scoring of impact?

Applicant characteristics do not appear to be associated with differences in scoring.

B3 – What factors affect impact scoring?

Since the introduction of the collection of more structured information under the Relative to Opportunity policy (R2O) and revised Statement of Expectations (SofE), Leadership level has become less predictive of the impact score. There is no apparent relationship between impact type and impact score.

Discussion

While noting the short time frames since RITRA was first implemented and that applicants and peer reviewers are still learning how best to engage with the RITRA framework, the evaluation has identified several issues arising from the framework implementation that need to be addressed. These include:

- duplication of information provided by applicants in the three impact subsections and between the impact and publication texts.
- difficultly experienced by applicants, and especially those early in their career, in providing retrospective impacts
- possible confusion about the nature of research impact
- difficulty experienced by peer reviewers in assessing the information provided in the applications
- insufficient guidance provided by the category descriptors.

Based on these findings, the RITRA framework could be improved to ensure it is achieving its short-term outcomes and overall objectives.

Recommendations

In order to address the issues identified by this evaluation, some revisions could be made to the RITRA framework, as follows:

- To reduce duplication of content in the application, the three components of the Research Impact section could be combined into one or two components
- To provide more space for applicants to explain their research impact, all evidence for impact – in the form of URLs or document citations – could be included within a separate free-text field of the application form
- To align with most applicants describing the generation of knowledge that may lead to impact, applicants could instead outline their pathway(s) to impact and engagement with research end users. NHMRC could develop advice about pathways to impact and describe various markers for each impact pathway type
- To help peer reviewers assess the applicant's impact pathway, the Category Descriptors could be revised – ideally assisted by cognitive interviews – to ensure that applicants, peer reviewers and NHMRC all understand them in the same way.

Revision of the RITRA framework should be guided by an expert working group.

Background

NHMRC's definition of impact

As set out within the Investigator Grants 2022 Guidelines, NHMRC defines 'Research impact' as:

> . . . the effect of the research after it has been adopted, adapted for use, or used to inform further research.
>
> - Research impact is the verifiable outcomes from research and not the prospective or anticipated effects of the research
> - Research impact also includes research that leads to a decision not to use a particular diagnostic, treatment or health policy.

For the purposes of RITRA implementation, NHMRC has recognised four types of research impact, being:

- Knowledge impact: new knowledge demonstrating the benefits emerging from adoption, adaption or use of new knowledge to inform further research, and/or understanding of what is effective.
- Health impact: improvements in health through new therapeutics, diagnostics, disease prevention or changes in behaviour; or improvements in disease prevention, diagnosis and treatment, management of health problems, health policy, health systems, and quality of life.
- Economic impact: improvements in the nation's economic performance through creation of new industries, jobs or valuable products, or reducing health care costs; improving efficiency in resource use; or improving the welfare/ well-being of the population within current health system resources. An economic impact may also contribute to social or health impacts, including human capital gains and the value of life and health.
- Social impact: improvements in the health of society, including the well-being of the end user and the community. This may include improved ability to access health care services; to participate socially (including empowerment and participation in decision making) and to quantify improvements in the health of society.

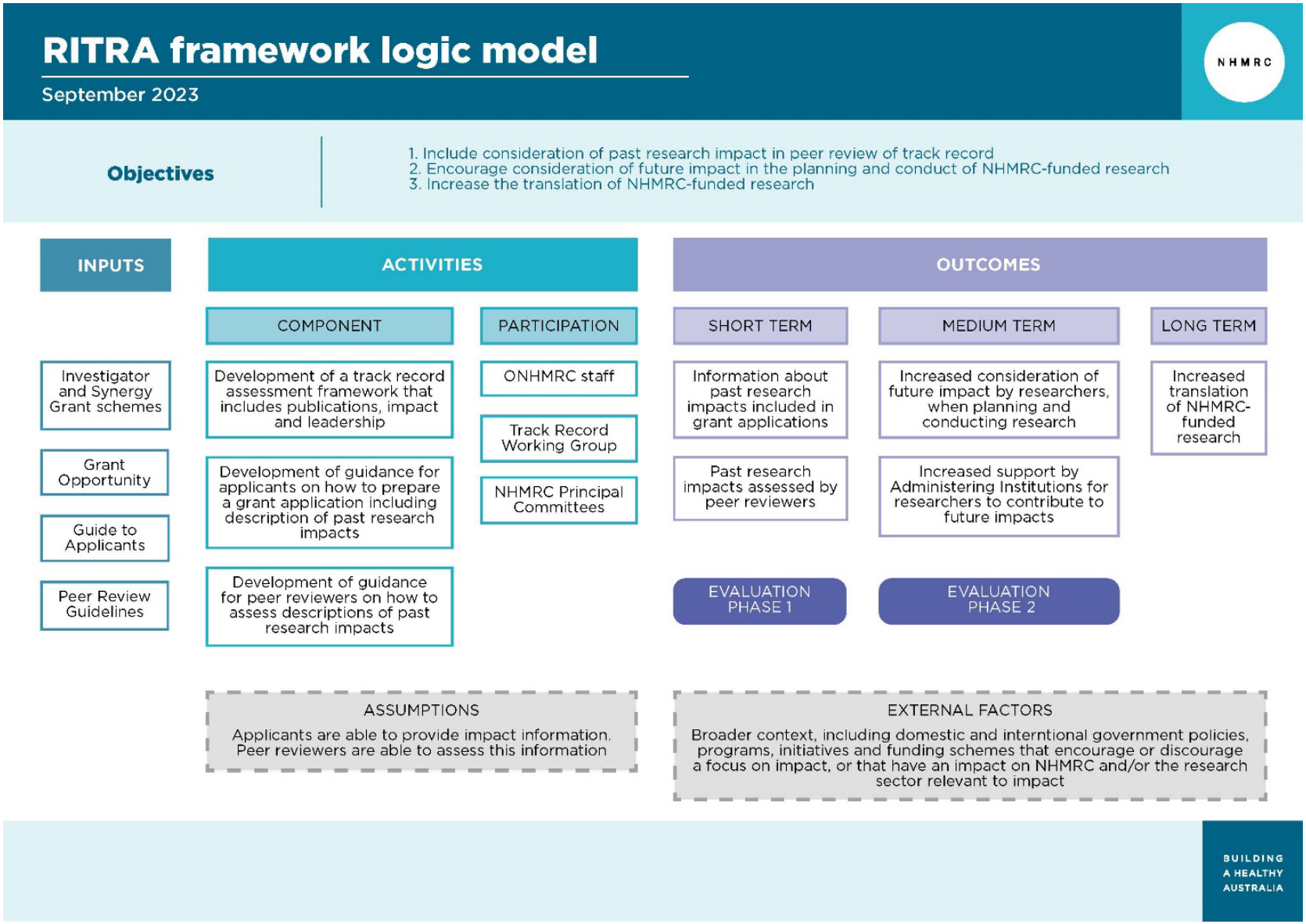
RITRA framework

NHMRC's implementation of the RITRA framework represents the agency's initial step towards increasing the translation of NHMRC-funded research through increasing the focus of NHMRC-funded researchers towards contributing to the delivery of research impacts.

As shown in Figure 1 (and also provided in text form in Appendix A), the framework has three key objectives. These are to:

1. include consideration of past research impact in peer review of track record (short-term)
2. encourage consideration of future impact in the planning and conduct of NHMRC-funded research (medium-term)
3. increase the translation of NHMRC-funded research (long-term).

Figure 1. RITRA framework logic model



Under the guidance of an Expert Advisory Group, development of options for a revised suite of NHMRC grant schemes commenced in 2016 and several working groups were established to facilitate this process, including a Track Record Working Group. This latter group developed a definition of track record: 'the value of an individual's past research achievements, not prospective achievements, quantitatively scored using evidence-based components. It also assisted in the development of a framework intended to shift the focus of grant application assessment away from traditional bibliometric measures of track record towards a greater emphasis on later stage outcomes and impacts.

The framework – which was incorporated in the Investigator Grant and Synergy Grant schemes commencing in 2019 – is composed of three major elements: publications, research impact and leadership.

Applications for Investigator Grants are assessed by peers against the assessment criteria of track record, relative to opportunity (70%) and knowledge gain (30%).

Assessment of track record in the Investigator Grant scheme comprises peer reviewers' consideration of:

- publications (35%)
- research impact (20%)
- leadership (15%).

Applicants are asked to demonstrate their research impact in three separate blocks of text, each a maximum of 3,000 characters, including any corroborating evidence. These blocks of text are:

Reach and significance: applicants are asked to describe the reach and significance of the research impact, including any corroborating evidence.

- Reach is the extent, spread, breadth, and/or diversity of the beneficiaries of the impact, relative to the type of research impact.
- Significance is the degree to which the impact has enabled, enriched, influenced, informed or changed the performance of policies, practices, products, services, culture, understanding, awareness or well-being of the beneficiaries (not the prevalence or magnitude of the issue).

Research program contribution: applicants are asked to describe their research program's contribution to the research impact.

- A research program is a cohesive body of research by the applicant. It is not limited to an individual case study (as used in a clinical context) or a single publication. A research program may be recent or in the past. Applicants need to outline the research program with corroborating evidence that can be independently assessed by peer reviewers.
- Research program's contribution to the research impact is the degree to which the applicant's research program was necessary to achieve the impact(s) (knowledge, health, economic, and/or social impact) based on robust and verifiable evidence. The relationship between the applicant's research program (including related activities) and the impact may be foreseen or unforeseen and may be an end-product or demonstrated during the research process. Research impact examples may include the adoption or adaptation of existing research.

Applicant contribution: applicants are asked to describe their contribution to the research program.

- An applicant's contribution to the research program is, relative to opportunity and to the applicant's field of research, the level of the applicant's contribution (e.g., leadership, intellectual and/or technical input) to the research program based on robust and verifiable evidence.

Peer reviewers score each block of text guided by the category descriptors shown in Table 1.

Applicants are provided with examples of evidence which are shown in Appendix B.

As shown in Table 1, different scoring is used for Emerging Leadership (EL) and Leadership (L) level applicants, in recognition of early career researchers (ECRs) having had less time to progress their research towards impact. EL1, EL2 and L grants are also funded from separate budgets, ensuring that ELs do not have to compete with Ls for funding.

Table 1. Category descriptors for research impact component

Emerging Leadership		Leadership	
Score	Description	Score	Description
7	an exceptional knowledge, health, economic and/or social impact (henceforth in this table, 'impact')	7 6 5	an exceptional impact an outstanding impact an excellent impact
6	an outstanding impact	4	a very good impact
5 4	an excellent impact a very good impact	3	a good impact
3 2	a good impact a satisfactory impact	2	a satisfactory impact
1	a weak or limited impact and/or the applicant has not supplied robust verifiable evidence	1	a weak or limited impact and/or the applicant has not supplied robust verifiable evidence

Evaluation of RITRA framework

Having now been used in over four rounds of the Investigator and Synergy Grant schemes. RITRA has had some time to influence researcher behaviour in grant applications, the behaviour of peer reviewers, and grant outcomes, but insufficient time to influence the way research is conducted, how research results are translated or the nature of research impacts.

Consequently, the present evaluation is a *process evaluation* undertaken to increase understanding of how effectively the implementation of RITRA within the Investigator Grant scheme has accomplished the framework's short-term objective.

Evaluation of the extent to which RITRA implementation has accomplished the framework's medium- and long-term objectives could be undertaken in the future, as required.

Evaluation questions

The questions that this evaluation seeks to answer are divided into two groups.

Group A questions examine the processes through which applicants provide impact text:

A1 – How easy was it to provide the impact text?
A2 – Are the impact types useful?
A3 – What types of impacts are being reported?
A4 – Is the impact text duplicating the publications text?

Group B questions examine the processes through which peer reviewers assess impact text:

B1 – How easy was it to assess the impact text?
B2 – Do applicant characteristics influence scoring of impact?
B3 – What factors affect impact scoring?

Rationales for these questions are provided in the 'background' sections below.

Methodology

To answer the questions listed above, the evaluation used a 'mixed methods' approach that included the following components.

Data analysis

The quantitative analysis for the evaluation was based on a dataset consisting of 6,755 applications to the NHMRC Investigator Grant scheme for the years 2019, 2020, 2021 and 2022 plus the scores provided by peer reviewers for each of the six assessment components, being:

1. publications
2. impact – Reach and significance (R+S)
3. impact – Research program contribution (RPCon)
4. impact – Applicant contribution (AppCon)
5. leadership
6. knowledge gain (for the project proposal).

While the dataset contains 6,755 applications, some of the fields within it contain missing or incomplete data. Therefore, some analyses of the whole dataset will refer to a total figure less than 6,755.

- Statistical techniques used include the calculation of correlations among variables and the use of linear regression to investigate how these variables predict total impact scoring.
  - A significance (p) value of <0.01 was used throughout the evaluation.
- Microsoft Excel, IBM SPSS and ProSuite WordStat were the software packages used to undertake this component.
- Descriptive statistics and charts for some of the fields included in the RITRA evaluation dataset are provided in Appendix C.

Surveys

Surveys, undertaken using Microsoft Forms, were conducted for 2023 Investigator Grants applicants and peer reviewers. The survey questions are provided in Appendix D and Appendix E.

Applicants for 2023 Investigator Grants were surveyed during the period 28 March to 14 April 2023. The survey was sent to 1,507 applicants and a total of 684 responses were received, providing a response rate of 45%.

Peer reviewers for 2023 Investigator Grants were surveyed during the period 26 July to 16 August 2023. The survey was sent to 500 peer reviewers and a total of 281 responses were received within this period, providing a response rate of 56.2%.[1]

Analysis of the survey results took the form of descriptive statistics plus a limited amount of qualitative analysis of the free text provided by survey respondents.

Where multiple choice components were left blank by less than 1% of respondents, these blank responses were excluded from the total number of responses for descriptive statistics. Where responses provided in free text fields were null text responses such as 'N/A', 'nil' or 'nothing to add', the analysis has assumed that these respondents did not realise these questions were optional and did not require any text input to move forward in the survey. Similarly, other comments such as 'I will come back to this later', 'see previous' or 'all good' did not contain meaningful feedback. To provide more accurate reporting data, these responses have been removed and classified as nil responses, so the total number of respondents reported in various parts of the analysis most accurately reflects those who provided substantive input.

Literature review

Both academic and other sources of information were consulted to provide context for various components of the evaluation.

The *Investigator Grants 2022 Guidelines* and *Investigator Grants 2022 Peer Review Guidelines* were used as primary sources of information about the Investigator Grant scheme, where relevant to the analysis.

Consultation

Members of NHMRC's Health Research Impact Committee (HRIC) Principal Committee were consulted about the RITRA evaluation at the committee's meetings in March, May and August 2022 and March and November 2023. Advice provided by committee members informed how the evaluation questions were framed, the types of analyses that were conducted and the report's final recommendations.

Evaluation

A1. How easy was it to provide the impact text?

Background

Reporting on impacts, and even conceptualising research outputs in terms of impacts, is a relatively new activity for NHMRC-funded researchers and something that, in other contexts[2]

[1] One late response was received. This response was not included in the analysis in this report.

[2] e.g. Bandola-Gill J, Smith KE. Governing by narratives: REF impact case studies and restrictive storytelling in performance measurement. Studies in Higher Education. 2022 Sep 2;47(9):1857-71

researchers have found difficult. Consequently, this task might prove challenging for some NHMRC grant applicants. It may also be the case that NHMRC's advice to applicants needs to be improved.

This question has been broken down into two sub-questions:

1. How easy was it for applicants to provide the impact text?
2. Could the way that NHMRC asks for impact text be improved?

Methodology

Applicant survey

Several questions in the survey of applicants addressed the questions above.

Applicants were asked to rate:

- the proposition, "The examples of evidence were helpful in supporting my impact case study statement within my application" (Q3)
  - applicants could also provide further comments (Q4)
- their ease of completing the three impact text components (Q5)
- the extent to which the information that they provided in these sections overlapped (Q6)
  - applicants could also provide further comments (Q7).

Peer reviewer survey

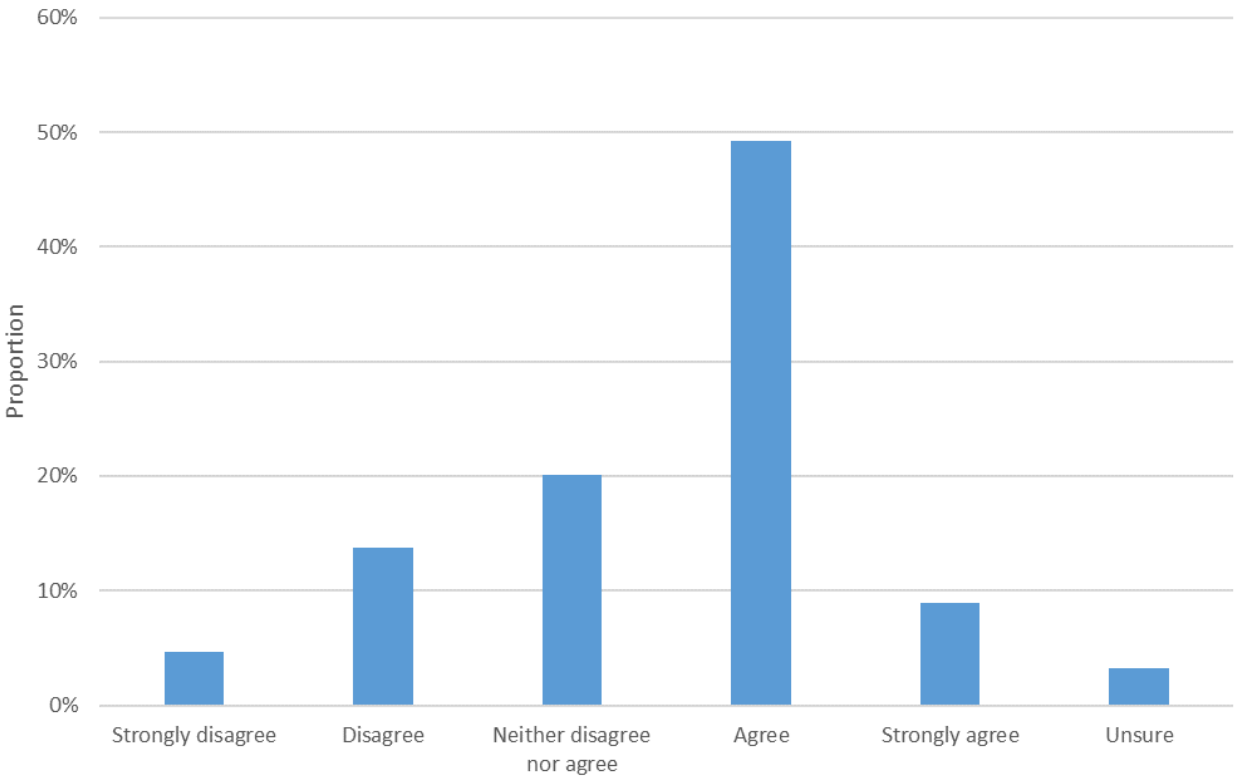
Peer reviewers were asked to rate how much the information provided by the majority of applicants within the three research impact components overlapped (Q25).

Results

Examples of evidence

As shown in Figure 2, 58% of applicants agreed or strongly agreed that the examples of evidence were helpful, while 19% disagreed or strongly disagreed, 20% neither disagreed nor agreed and 3% were unsure.

Figure 2. **Applicant responses to Q3: "The examples of evidence were helpful in supporting the research impact components within my application."**



Of the respondents, 30.8% (211 of 684) provided a free text comment about this issue, and 93.8% of these (198 of 211) flagged concerns and provided suggestions for consideration by NHMRC. These included:

- improving the examples of evidence to reduce ambiguity and confusion by applicants
- providing more examples of evidence and more guidance on how to demonstrate impact using these examples
- extending the character count to provide sufficient space for examples of evidence
  - examples of evidence could be included in a separate section that does not contribute to the character count. As references and weblinks vary in character count, this would ensure that applicants are not unduly disadvantaged when including examples of evidence of higher character count
- developing a template or spreadsheet for applicants to log impact and gather evidence, which could then be referred to in the future as applicants draft their applications
- providing more specific advice, examples and information to assist ECRs in providing examples of evidence where their work may not have yet generated a high level of impact
- allowing applicants to use emails as examples of evidence
- allowing applicants to use web address shorteners to save character space
- improving the language used to describe the examples of evidence (i.e., current language is corporate and not reflective of science research and outcomes)

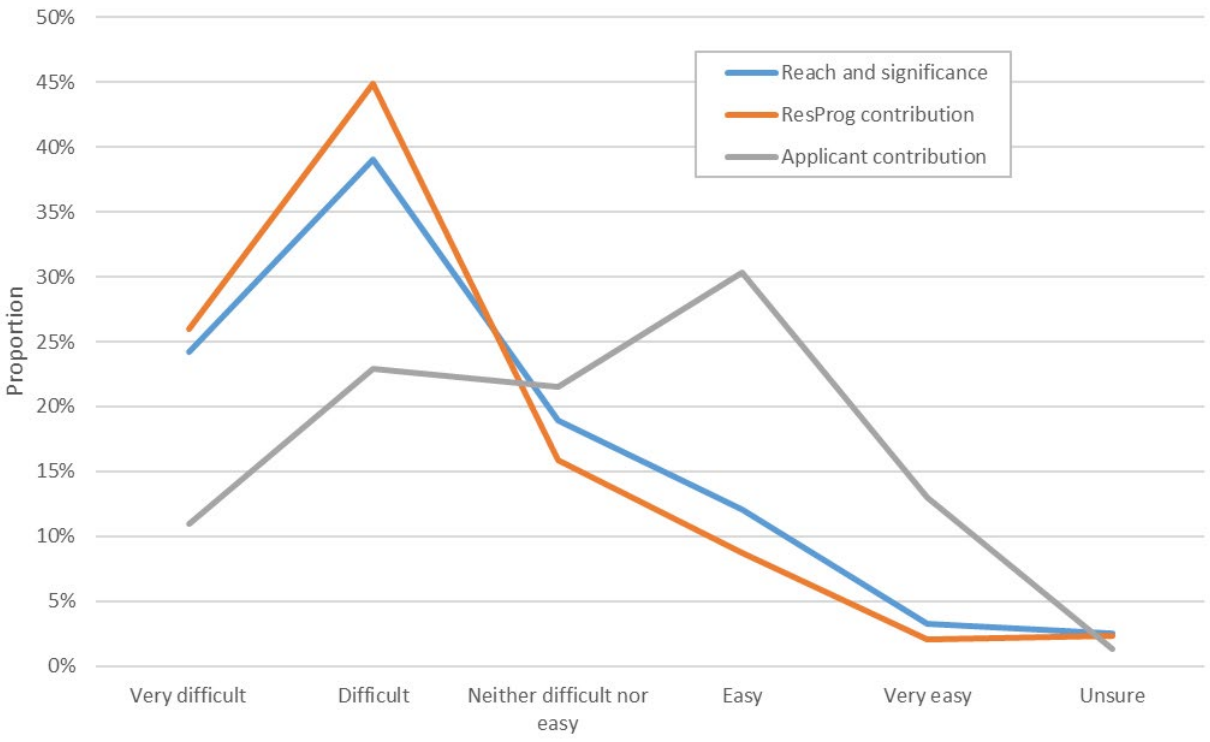
- NHMRC should provide clarity on whether more or fewer examples of evidence are beneficial for an applicant. "When a list is provided, it also inherently implies that you need to tick off as many as possible, which dilutes the amount of words you can dedicate to each. If this is not the intention, this should be made clear."
- providing guidance on what evidence should be provided in which of the three sections
- reviewing and updating the list of examples of evidence annually based on successful Investigator Grant applications.

In addition, several applicants noted that private sector businesses existed that provided in-depth advice on examples of impact that is clearer and provided them with a better understanding of impact compared to the advice provided by NHMRC.

Completing the impact sections

As shown in Figure 3, applicants found the R+S and RPCon sections more difficult to complete than the AppCon section.

Figure 3. Applicant responses to Q5: 'Filling in the following components is . . ."



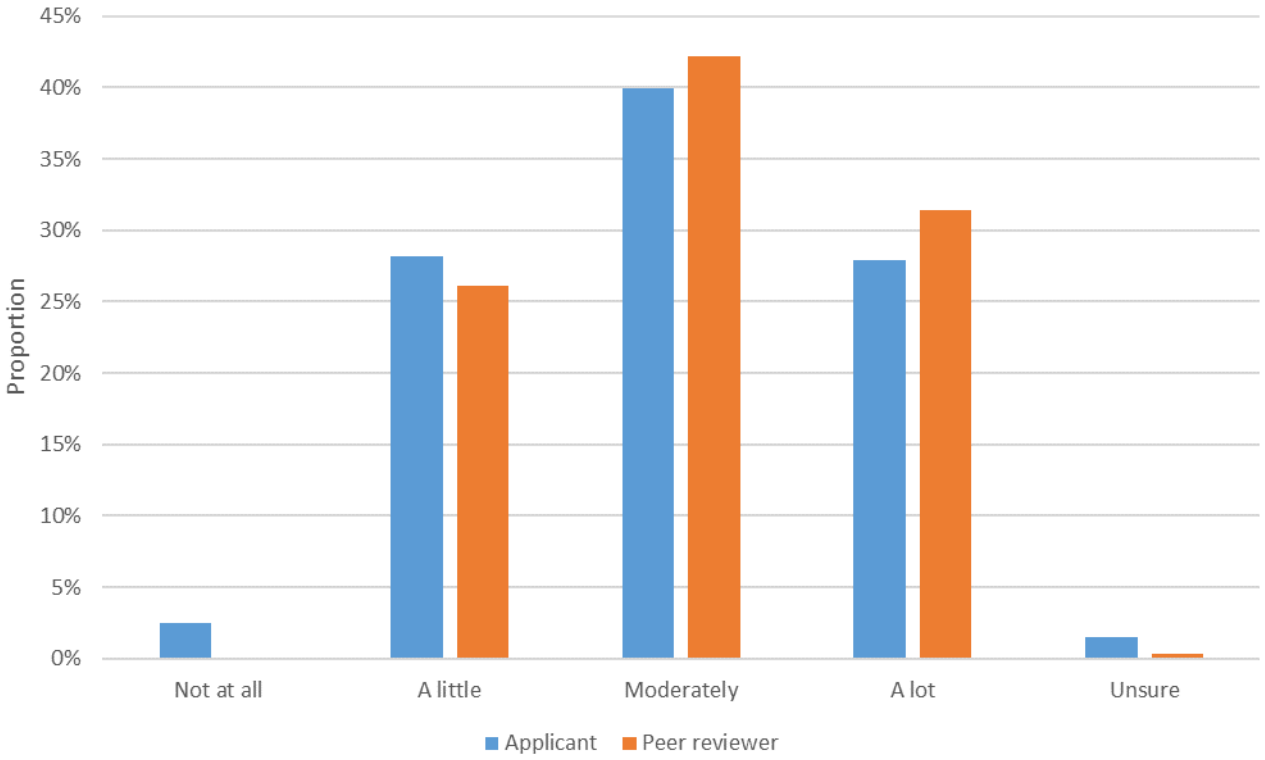
On average, 67% of applicants found the R+S and RPCon sections difficult or very difficult to complete, compared with only 34% of applicants for the AppCon section.

Conversely, on average only 13% of applicants found the R+S and RPCon sections easy or very easy to complete, compared with 43% of applicants for the AppCon section.

Furthermore, and as shown in Figure 4:

- 68% of applicant respondents thought that the information that they provided for the three sections overlapped moderately or a lot, while 31% of applicant respondents thought that there was little or no overlap.
- 74% of peer reviewers thought that the information provided by the majority of applicants within the three impact components overlapped moderately or a lot, while 26% of peer reviewers thought that the texts overlapped little. No peer reviewers indicated that the text overlapped 'not at all'.

Figure 4. **Applicant and peer reviewer responses to Q6 and Q25 (respectively) on overlap in information provided for the three impact components**



Applicant comments

Of the respondents, 60.4% (413 of 684) provided a comment in the free text field (Q7) relevant to Q5 and Q6. Most respondents offered suggestions for improvements across the three free-text field components.

The comments discussed:

- Overlap between the three sections (31.2% of respondents), indicating this is the most significant issue faced by applicants. On this topic, applicants detail overlaps across all three sections or two sections only (most often R+S and RPCon; or RPCon and AppCon).
- How to reduce the overlap of information in the three impact components. Comments and suggestions included:
  - 43 comments (10.4%) recommended combining components, that is, either combining all three components (into one overall section that allows applicants to detail their impact(s) in

a manner that suits their research story) or combining two of the three components (to provide applicants with two distinct sections to minimise overlap).

  - allowing applicants an additional area at the start of the research impact section where they can detail their research program

- 28 respondents (6.8%) suggested that NHMRC change the order of the components. Currently, applicants are required to fill out R+S first, followed by RPCon and then AppCon. Most comments suggest that, following the order of what occurs in a research project, these components should be reversed. This would assist applicants to reduce overlap across the three components and provide clarity between an applicant's contribution and the research program's contribution.
- 32 respondents (7.7%) discussed their need to engage with or acquire external advice to clarify the aspects of the research impact section. External advice included the use of senior researchers who had successfully obtained NHMRC funding in previous grant rounds, seminars on writing research impact statements, Research Administration Officers, university grant officers and paid external feedback consultants.
- 22 respondents (5.3%) discussed increasing the character limit for the components. A subset of these respondents added that providing evidence, particularly URLs (which can vary in length) has a negative impact on the number of characters left to describe their research impact.
- A smaller number of respondents (5) described the research impact section as a narrative that benefits those who are able to self-promote or market themselves effectively, rather than a true reflection of a researcher's impact across their career. This may negatively impact applicants who experience difficulties with written communication, or those for whom English is not their primary language.
- Another subset of respondents (8) provided details of the research impact section being easy to write as a clinical researcher but more difficult for basic scientists. These commentors suggested that the section should be reworked to provide components that are fairer to scientists from all sectors and backgrounds.

Other common topics of discussion included assessing whole-of-career impact, providing clearer components and definitions (using plain English) and providing better examples and guidance. A number of respondents also suggested providing different components for different impact types, or different components for different application levels, so they can most accurately reflect an applicant's research career level.

Some applicants have reported concern over the number of changes that have occurred in the grant application process over the past ten years. Some respondents have indicated that these changes require time to understand and multiple grant applications to feel as though they are 'getting it right'. If NHMRC is to change these processes once again, it should ensure the change is driven and informed by the sector.

Findings

While applicants found the examples of evidence useful, they found it difficult to provide impact text about the research program's contribution to the impact, and about the reach and significance of the research. Comparatively, most applicants found their own contribution easy or very easy to write about.

Both applicants and peer reviewers thought that there was substantial overlap in the content of the three impact text sections.

A2 – Are the impact types useful?

Background

While numerous extra-academic benefits may arise from research, for ease of discussion and comprehension it has been common for both funders and research organisations to develop lists of impact categories or types. For example, University College Dublin proposes the following as 'types of impact':[3]

- cultural
- economic
- environmental
- health
- political
- scientific
- social
- technological
- training.

CQUniversity's list[4] is similar but not identical, consisting of:

- academic impact
- cultural impact
- economic impact
- wellbeing impact
- policy impact
- environmental impact
- social impact
- training impact

Impact may also be viewed from the perspectives of particular groups of stakeholders. For example, Bainbridge et al 2015[5] state:

> In Aboriginal and Torres Strait Islander health research, research benefit is broadly defined as any elements of research that are advantageous or good; such as strengthening capacities, opening opportunities or improving health outcomes that progress the interests that are valued by Aboriginal and Torres Strait Islander (respectfully hereafter Indigenous) people . . .
>
> In a policy context, Indigenous Australian people define research benefit as, "the establishment or enhancement of capacities, opportunities or outcomes that advance

[3] Refer: https://www.ucd.ie/research/portal/typesofimpact/

[4] Refer: https://libguides.library.cqu.edu.au/c.php?g=881139&p=6659585

[5] Bainbridge R, Tsey K, McCalman J, Kinchin I, Saunders V, Watkin Lui F, Cadet-James Y, Miller A and Lawson K. No one's discussing the elephant in the room: contemplating questions of research impact and benefit in Aboriginal and Torres Strait Islander Australian health research. BMC Public Health. 2015 Dec;15:1-0

the interests of Aboriginal and Torres Strait Islander peoples and that are valued by them."

Because of the multiplicity of both potential types of impact and potential beneficiaries, there can be no ideal or final list of impacts, and the development of any actual list will be subject to the details of time, location and stakeholders.

One consequence of the lack of an authoritative impact type list is that NHMRC's impact typology may not best meet the needs of Investigator Grant scheme applicants or peer reviewers.

Questions:

1. Are the impact types indicated by applicants a reliable source of information about the content of their impact text?
2. Do applicants benefit from framing their responses in terms of impact types?

Methodology

Data analysis

Question 1, above, was addressed by examining the presence of keywords within the impact text provided by each applicant.

A list of these keywords for each impact type was extracted from the examples of evidence tables within the Investigator Grants 2022 Guidelines (refer Appendix B) and was supplemented by looking for synonyms of these words within the total word frequency list generated for all the impact text. The keyword list (Table 2) is not considered to be exhaustive but did not need to be for the purposes of this analysis, which was to determine whether applicants who indicated a particular impact type included a larger number of the related keywords than applicants who did not.

The keywords were searched for within each applicant's total impact text. Word stems, rather than whole words, were searched for to reduce the number of whole-word variants searched for.

Table 2. **Impact type keyword list**

Impact type	Keywords
Knowledge impact	academ*, article, author, cita*, cite, conference, data, doi, edit*, fwci, google, index, invit*, journal, keynote, knowledge, literature, manuscript, pmid, presentat*, prize, publica*, publish, scival, scopus, weighted
Health impact	clinic, cochrane, diagnos*, therap*, guideline, health, hospital, medic*, patient, random, rct, practice, manage, prevent, treat, quality, phase
Social impact	social, society, communit*, decisionenvironment, govern*, polic*, risk, service, determinant, inclusion, maker, particip*
Economic impact	commerc*, company, consumer, device, drug, econom*, engineer, fda,incorporate, industry, ip, patent, PCT, pharma*, technol*, employ*, cost, regulator

Surveys

Question 2 above was addressed through survey questions.

Applicants were asked to rate their agreement with the statement, "The four impact types allow me to report all the impacts I would like to" (Q1).

- They could also "Provide any comments on the definitions of the four impact types and/or any other impact types you would like included" (Q2).

Peer reviewers were asked to rate their agreement with the statement, "The four impact types cover the impacts described by applicants" (Q21).

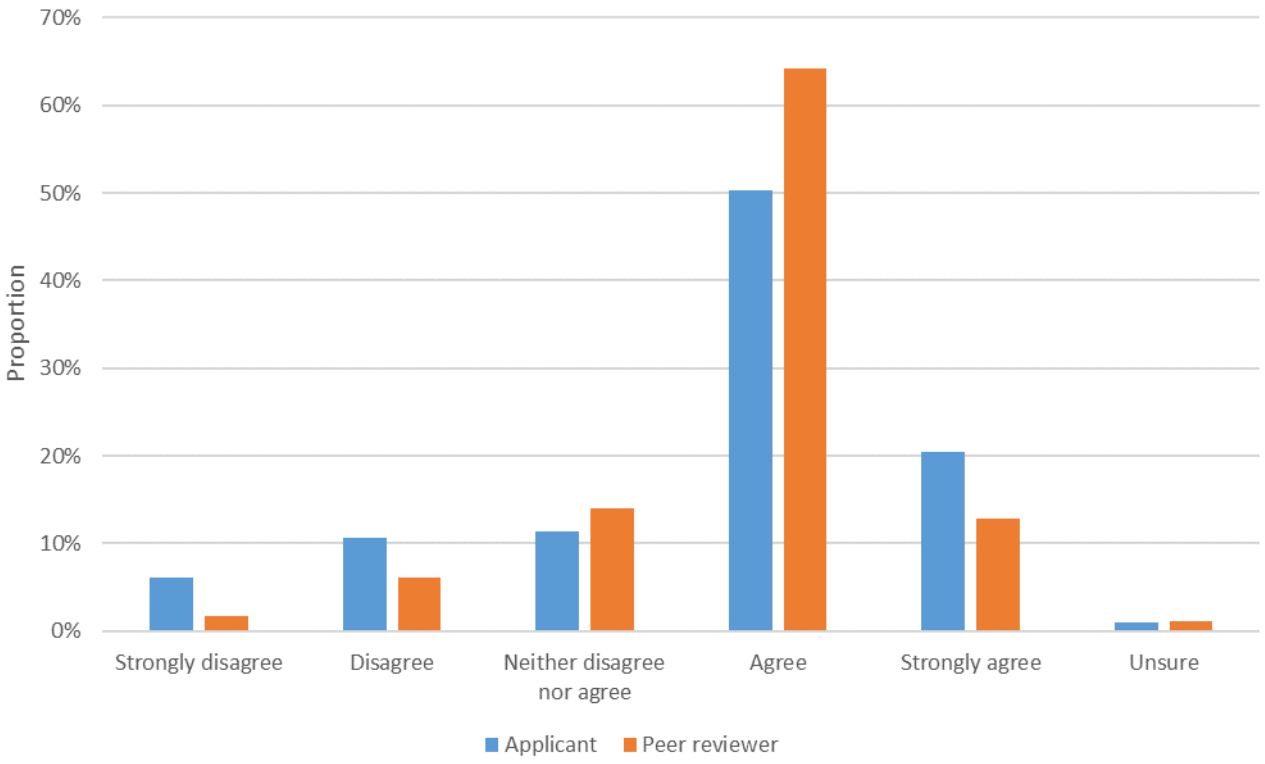
- Peer reviewers could also "Provide any comments on the definitions of the four impact types and/or any other impact types you would like included" (Q23).

Results

Applicant survey

As shown in Figure 5, 70% of applicants agreed or strongly agreed that the four impact types allowed them to report all the impacts that they wished to, while 17% strongly disagreed or disagreed, 11% neither disagreed or agreed, and 1% were unsure.

Figure 5. Applicant and peer reviewer responses to Q1 and Q21 (respectively) about the suitability of the four impact types



Applicants were asked to provide any comments on, "the definitions of the four impact types and/or any other impact types you would like included".

Of the respondents, 38.9% (266 of 684) provided a comment in response to this question. These responses were qualitatively grouped into positive, neutral and negative by analysing the key themes and tone of the responses.

From the respondents, 9.4% (25 of 266) provided positive feedback, i.e., indicating that the definitions of the four impact types were adequate, appropriate and covered all impact types. As such, the remaining 90.6% (241 of 266) of comments were classified as neutral or negative.

Respondents flagged several concerns and provided feedback for consideration by NHMRC. These included:

- 23.7% (63) of responses indicated that impact is difficult to characterise and the segregation into specific impact types is not useful. Overwhelmingly, respondents suggested that there should be no selection of impact type, and instead, applicants should be required to describe their overall impact without needing to differentiate between types.
- 10.5% (38) of respondents discussed the notion of significant overlap between the impact types currently provided by NHMRC, particularly knowledge and health impacts. Some respondents suggested that NHMRC provide clearer distinctions between the impact types.
- Some respondents were concerned that some impact types might be being scored more favourably by peer reviewers than others and their mistrust or unease with the effect this may have on the favourable scoring of their impact sections.
- Respondents also indicated that they were unsure if there were advantages or disadvantages to selecting more than one impact, even where their research falls over several impact types. Several respondents stated that they were informed that they would be disadvantaged by selecting two or more impact type and had been advised by their research institutes to select only one.
- Respondents suggested the inclusion of additional impact types. Where these impacts (or outcomes) cannot be included in the list, it was suggested that NHMRC provide guidance about which of the four impact types they would fit under. Suggested impacts include:
  - educational impact
  - Aboriginal and Torres Strait Islander research impact
  - equity impact
  - infrastructure impact
  - innovation impact
  - policy impact (including government committees, guidelines, reports)
  - sector training impact
  - technology advancement impact
  - health promotion impact (including non-academic publications and media)
  - community engagement.
- A small number of respondents commented on the disadvantages faced by ECRs in relation to impact types. They discussed that, for ECRs, it is difficult to demonstrate impact early in their career and as such, they are at a disadvantage under these impact types.
- Other common themes included the notion that the character limit is restrictive and should be increased and that the impact definitions and examples are too broad.

Peer reviewer survey

Also as shown in Figure 5, peer reviewers generally agreed that the four impact types covered the applicants' described impacts.

Of the peer reviewer respondents 19.9% (56 of 281) provided a comment in response to the prompt, "provide any comments on the definitions of the four impact types and/or any other impact types you would like included" (Q23).

Common comments by peer reviewers (% of respondents) included:

- the presence of overlap or repetition between impact sections or impact types (12.5%)
- that applicants often chose the wrong impact type (8.9%) which did not match with the impacts and evidence described
- that applicants had difficulty distinguishing impact categories (7.1%) which was reflected in applicant responses to the research impact track record section
- that most applicants selected knowledge impact (8.9%)
- that economic impact and social impact were the least commonly reported impact types (8.9%).

Peer reviewer respondents (% of respondents) suggested the following changes:

- removal of impact type selection (14.3%)
- providing better and clearer guidance to applicants by:
  - improving and expanding definitions of impact types, particularly social impact and the distinction between health and knowledge impacts
  - providing a "perfect example" for applicants
  - providing specific examples relevant for EL applicants
  - improving guidance of what information belongs in which response section.
- provision of more guidance to peer reviewers on how to assess this section.

Data analysis

As shown in Table 3, applications that indicated a particular impact type included more keywords relevant to that impact type than those that did not. This was most obviously true for economic impact, where 517 applications selected this impact type and included almost 10,000 relevant keywords, or 18.8 per application. By contrast, the 6,236 applications that did not select economic impact included 40,383 economic terms, or only 6.5 per application. The ratio of keywords per application of those applications that did and did not select economic as an impact type is 18.8/6.5=2.9.

As shown in Table 3, the ratio of keywords per application that did versus did not indicate each impact type is, from largest to smallest: 2.9 (Economic), 2.7 (Social), 1.8 (Health) and 1.5 (Knowledge).

Whether or not an applicant indicated knowledge impact made little difference to the number of knowledge impact-related terms that they included in their impact text. The 593 applicants who did not indicate knowledge impact still included almost 13,000 knowledge impact terms (or about 22 terms per application), which is more than the number of social terms per application included by social impact applicants (16.6) and more than the number of economic terms per application included by economic impact applicants (18.8).

Table 3. **Impact types versus Impact-type related terms**

Impact type	Type selected	Sum of keywords	Count of applications	% of applications	Keywords/ application	Ratio of 1/0
Knowledge	0	12,822	593	8.8%	21.6	1.5
	1	204,649	6,160	91.2%	33.2	
Health	0	82,979	4,056	60.1%	20.5	1.8
	1	98,776	2,697	39.9%	36.6	
Social	0	38,873	6,254	92.6%	6.2	2.7
	1	8,268	499	7.4%	16.6	
Economic	0	40,383	6,236	92.3%	6.5	2.9
	1	9,696	517	7.7%	18.8	

Another way to determine whether impact types are related to differences in impact text is to look at word frequency – or more precisely, term frequency–inverse document frequency (tf-idf),[6] – across the impact types. Table 4 shows the 40 words with the highest tf-idf scores within the impact text for those applicants who selected each impact type. As noted above, there is considerable overlap between these groups, since 91% of applicants selected knowledge and 40% selected health.

In Table 4, grey shading indicates that a word was present in more than one column while yellow shading and a # indicates that a word was present in only one column.

Table 4. **Top 40 words by tf-idf score by impact type**

#	Knowledge	Health	Social	Economic
1	AL*	AL*	MENTAL	#ECONOMIC
2	PMID	CARE	HIV	CANCER
3	CANCER	CANCER	#ABORIGINAL	DRUG
4	CELL	PMID	POLICY	#TECHNOLOGY
5	CITES	HEALTH	#SOCIAL	#PATENT
6	FWCI	GUIDELINES	HEALTH	#WO*
7	#BRAIN	HIV	CARE	AL*
8	CARE	HTTPS	ALCOHOL	#DEVICE
9	HTTPS	TRIAL	NSW	PMID
10	PAIN	PATIENTS	#SUICIDE	CELL
11	HIV	MENTAL	PREVENTION	#PATENTS

[6] This statistic reflects how important a word is to a document or block of text that is itself part of a larger collection or corpus. The tf–idf value increases proportionally to the number of times a word appears in the document/text block and is offset by the number of documents/text blocks in the corpus that contain the word, which adjusts for some words appearing more frequently in general. Higher values of this statistic indicate that a term is more important or relevant to a specific body of text.

12	CELLS	RISK	INDIGENOUS	#INDUSTRY
13	HEALTH	POLICY	PEOPLE	#COMPANY
14	RISK	#PRACTICE	#SERVICES	CITES
15	#MENTAL	TRIALS	AL*	#COST
16	DRUG	#TREATMENT	#COMMUNITY	TB
17	#DIABETES	FWCI	CHILDREN	CARE
18	TRIAL	CHILDREN	#CHILD	TRIALS
19	STROKE	WWW	#GOVERNMENT	PATIENTS
20	DISEASE	#STROKE	#DEMENTIA	#COMMERCIAL
21	PATIENTS	CITES	HTTPS	VACCINE
22	#NAT*	NSW	#REPORT	TRIAL
23	#NATURE	PREVENTION	#SERVICE	#DELIVERY
24	#GENETIC	CLINICAL	#PUBLIC	HEALTH
25	TRIALS	ALCOHOL	#MEN	STROKE
26	#BONE	AU	CANCER	TREATMENT
27	#EXERCISE	#MANAGEMENT	#COMMUNITIES	#PHASE
28	GUIDELINES	PAIN	AU	IMAGING
29	HEART	#EXERCISE	#FOOD	CLINICAL
30	TREATMENT	#LANCET	SLEEP	CELLS
31	CHILDREN	INDIGENOUS	#EVALUATION	NSW
32	WWW	#SERVICES	#DISABILITY	#MALARIA
33	VACCINE	#PATIENT	#INTERVENTIONS	#MEDICAL
34	#PROF	PEOPLE	GUIDELINES	AU
35	SLEEP	#AF*	#EDUCATION	HTTPS
36	#MED*	TB	WWW	DISEASE
37	CLINICAL	#WOMEN	IMPLEMENTATION	#CHEM*
38	IMAGING	HEART	#INTERVENTION	HEART
39	POLICY	IMPLEMENTATION	#COMMISSION	#COMPANIES
40	HUMAN	SLEEP	#SAFETY	#MATERIALS

Some of these 'words' are actually abbreviations commonly used in the impact text.

- 'AL' is found within 'et al', as part of journal references.
- 'NAT', 'MED' and 'CHEM' are used in journal titles.
- "HTTPS", "WWW" and "AU" are used in URLs.
- The two-letter code "WO" refers to a publication from a PCT application (Patent Cooperation Treaty), which is the international patent system and stands for Written Opinion.
- "AF" stands for atrial fibrillation.

As may be seen in Table 4, while there are many words that occur within the top 40 most frequently used words for all impact types, there are also some that are unique to the top 40 of each impact type. This is most true for the Social impact type, where 21 of the top 40 words are unique to Social impact, and also true for the Economic impact type, where 17 of the top 40 words are unique to Economic impact. Each of these two impact types, of course, has a small overlap with the other types, unlike Knowledge impact, which (as noted above) was selected by 91% of applicants.

Findings

The survey data demonstrates that the majority of applicants consider that the four impact types allow them to report all the impacts that they would like to.

Peer reviewers report that the impact types provided to applicants covered the impacts described by applicants, however some raised concerns that applicants chose the wrong impact type and had trouble distinguishing impact types.

The data analysis confirms that the impact type being selected by applicants does relate to the type of information that they include within their impact text and that this effect is strongest for the least used impact categories (social and economic).

Conversely, the analysis also reveals that the knowledge impact type is least useful: whether or not an applicant selected knowledge impact made little difference to the number of knowledge impact-related terms that they included in their impact text.

A3 – What types of impacts are being reported?

Background

Historically, when researchers have used the term 'impact' to describe the outcomes of research, their focus has been on the impact of that research on knowledge.

As noted by Deeming, Reeves, Ramanathan et al 2018[7], when referring to the results of a survey that they undertook as part of their investigations of the attitudes and opinions of Australian medical research institutes towards research impact assessment frameworks:

> Albeit acknowledged as a legacy interpretation, research impact was presented by some as the traditional academic measures of publications, citations, other bibliometrics, grants and awards. The more contemporary interpretation of research impact reflected the challenge "to change lives", extend lifespans and improve quality of life.

Similarly, Penfield et al 2014[8] state:

> From the outset, we note that the understanding of the term impact differs between users and audiences. There is a distinction between 'academic impact' understood as

[7] Deeming S, Reeves P, Ramanathan S et al. Measuring research impact in medical research institutes: a qualitative study of the attitudes and opinions of Australian medical research institutes towards research impact assessment frameworks. Health Research Policy and Systems, 2018 16, 28

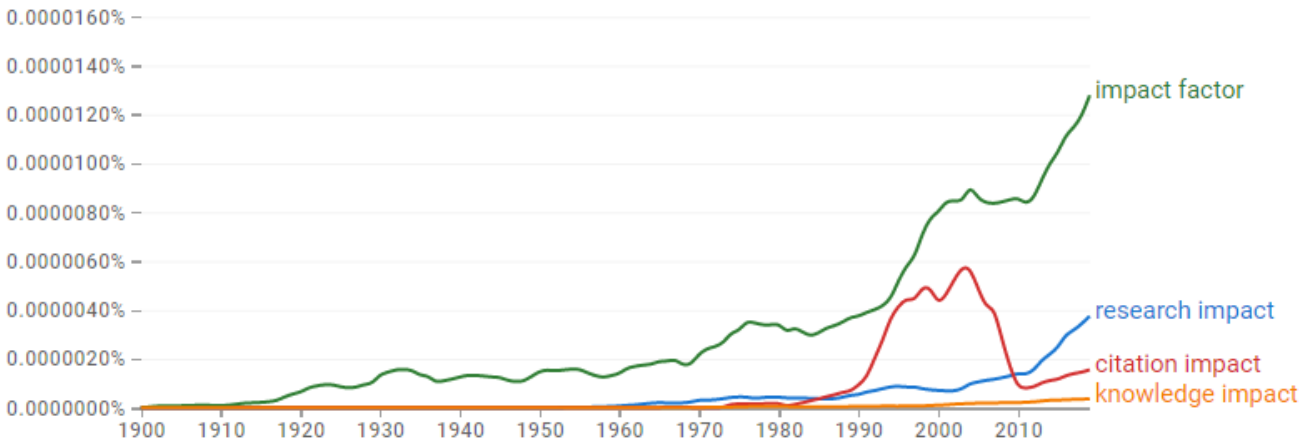
[8] Penfield T, Baker MJ, Scoble R and Wykes MC. Assessment, evaluations, and definitions of research impact: A review. Research evaluation. 2014 Jan 1;23(1):21-32

the intellectual contribution to one's field of study within academia and 'external socioeconomic impact' beyond academia.

Terms containing the word impact that are commonly used within the research sector include 'impact factor' and 'citation impact'. Changes to the usage of these terms over time are depicted in Figure 6.

The term 'impact factor' as it applies to journals (i.e. 'journal impact factor') dates from 1961 and thus the presence of this term in Figure 6 before that date must relate to its use in other contexts. However, even when using the 1961 level as a baseline it is clear that 'impact factor' has always been in greater use than the term 'research impact'. 'Knowledge impact' is a term that is of recent origin and is not much used.

Figure 6. Percentage of books that phrases occurred in, by year (1900-2019) by phrase[9]



Another way of discovering the different frequencies of use of these expressions is to look at the number of pageviews received by relevant Wikipedia pages (refer Figure 7). For the period 7 January 2015 to 22 March 2023, these are:

Research impact – 868

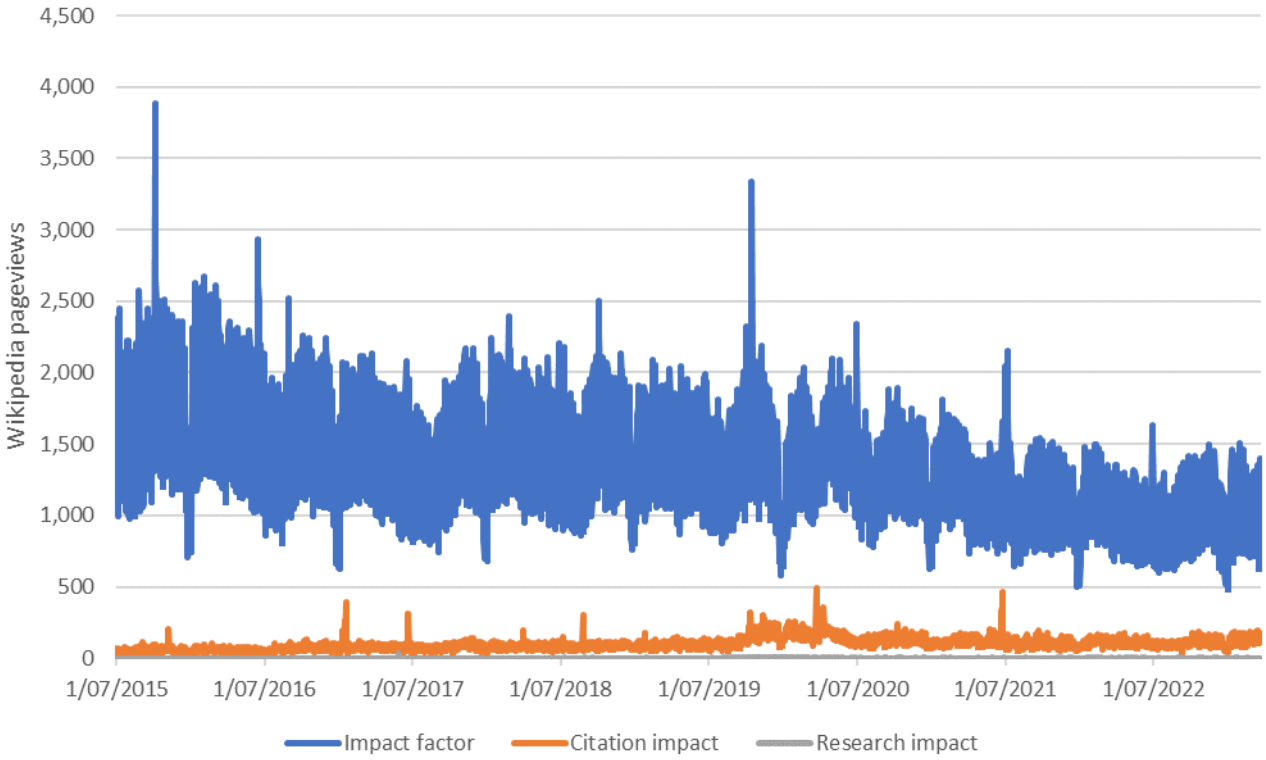
Citation impact – 287,877

Impact factor – 4,100,359.[10]

[9] Refer: https://books.google.com/ngrams/info

[10] Data available from: https://pageviews.wmcloud.org/

Figure 7. Number of pageviews received by Wikipedia pages (7/1/15-22/3/23), by page name



The term 'impact factor' has been used more than the term 'research impact' because the former is of most interest to researchers (a numerous group) while the latter is of most interest to research funders (a much smaller group).

When funders such as governments and philanthropic organisations support research they do so in order to generate benefits for a group of stakeholders (e.g. the public), so for funders, 'research impact' equates to 'stakeholder benefits'.

In the large majority of cases, however, stakeholder benefits are not something that researchers are able to produce. For most researchers, engaging with potential end-users of their research is the most that is possible. For this reason, funder use of the term 'research impact' can be confusing for researchers.

As noted by the University of York:[11]

> It can sometimes be hard to distinguish the difference between the routes to research impact (often known as knowledge exchange, engagement, or impact activities) and the impact itself. You may find it helpful to consider impact as something that other people or institutions gain or do - it is not something that you as a researcher can 'do', rather you can encourage it through impact activities. For example: giving evidence to a select committee, working with a business or contributing to an exhibition in a museum is not impact. It would become impact if the people or organisation involved somehow do, experience or understand differently as a result of their interaction with the research.

[11] Refer https://www.york.ac.uk/staff/research/research-impact/impact-definition/

Because decades may elapse between when research takes place and when impacts arising from it become apparent, only some – and usually senior – researchers will be able to make claims about health, social or economic impacts arising from their research. Most others will be limited to reporting on progress that has been made along a pathway between their research and possible future impacts.

In harmony with a focus on 'pathways to impact', the definition of research impact used by the Australian Research Council is the "contribution that research makes to the economy, society, environment or culture, *beyond the contribution to academic research*" (italics added).

According to this definition, contributions to academic research are not research impacts, but nonetheless they are key events on an impact pathway.

Question:

1. Do the types of impact claims being made by Investigator Grant applicants within the impact text focus more on pathways to impact (as might be predicted) or more on the direct benefits experienced by stakeholders?

Methodology

Studying the impact research and significance text is challenging because it is free text, and applicants have reported in a variety of ways. That said, given that the database contains over 6,700 reach and significance statements[12], it is possible to look for patterns within them. This can be done in at least two ways:

1. Substring search: this involves breaking each statement into its component sentences (substrings) and then looking for common types of impact claim text within them.
2. Factor analysis: this is a statistical technique that is used to reduce many variables into a smaller number of factors. This technique extracts maximum common variance from all variables and puts them into a common score.

Results

Substring search

Using Excel to divide each reach and significance 'string' into a collection of sub-strings divided by full stops (periods) led to the generation of 22,951 separate substrings. Not all of these were whole sentences as periods are used for a variety of purposes besides terminating a sentence. That said, both whole and partial sentence substrings were analysed for their content.

These substrings were reviewed to determine whether any patterns in the text were associated with impact claims. Two types of patterns were readily apparent: applicants would indicate impacts by using the words 'new' and 'novel', and by reporting that their research 'led to' some outcome.

[12] The dataset contains multiple applications (to different rounds) from some researchers and some of these statements contain similar blocks of text. That said, for the purposes of this evaluation each application has been treated as unique.

Of the 1,919 substrings that included the word ‘new’, as shown in Figure 8, 334 listed ‘new knowledge’. All of the other options were much less common. ‘New therapeutic’ was mentioned 71 times; ‘new field’ was mentioned 62 times.

Figure 8. Count of most commonly reported ‘new’ developments

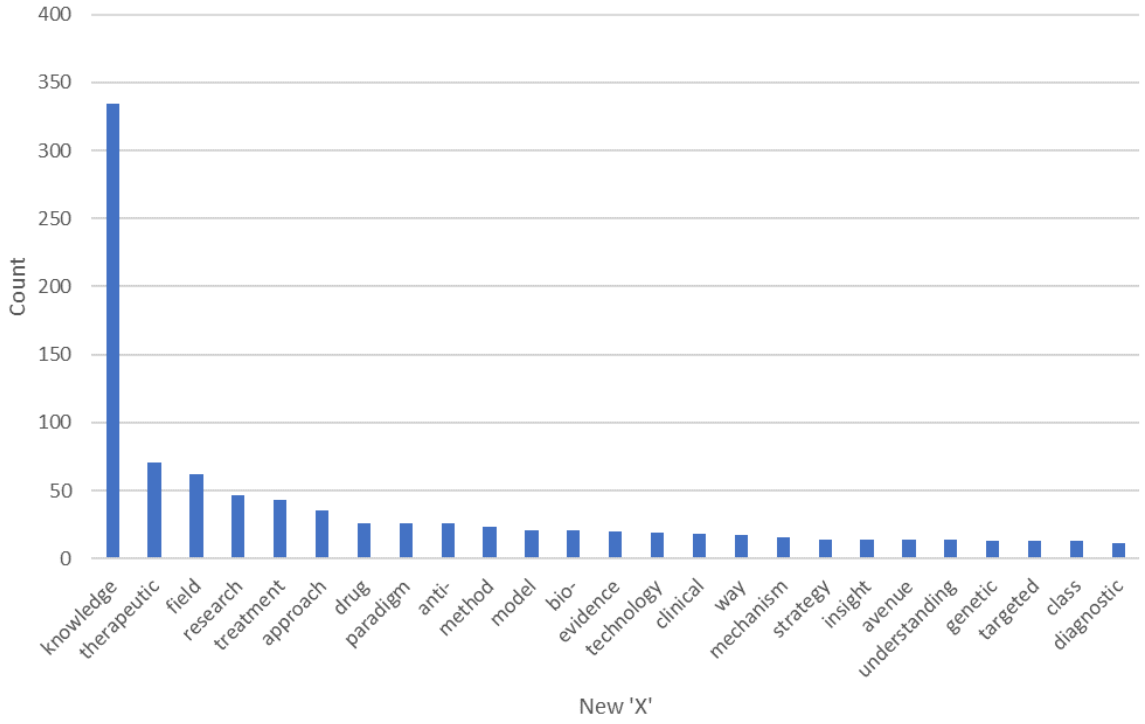
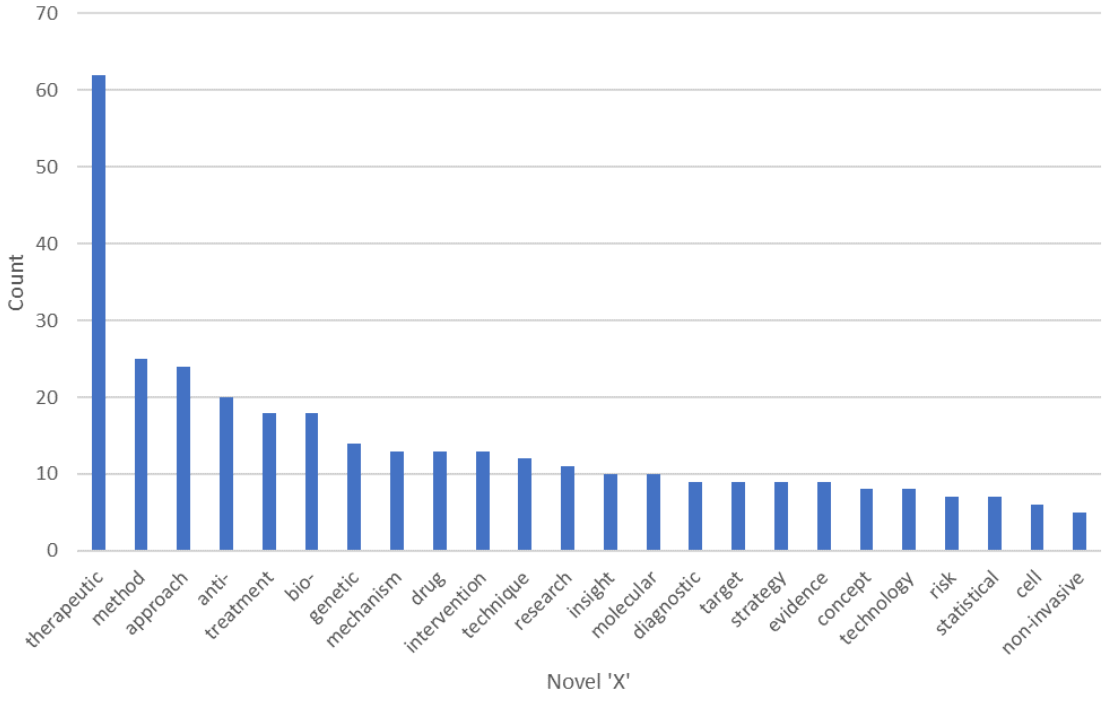


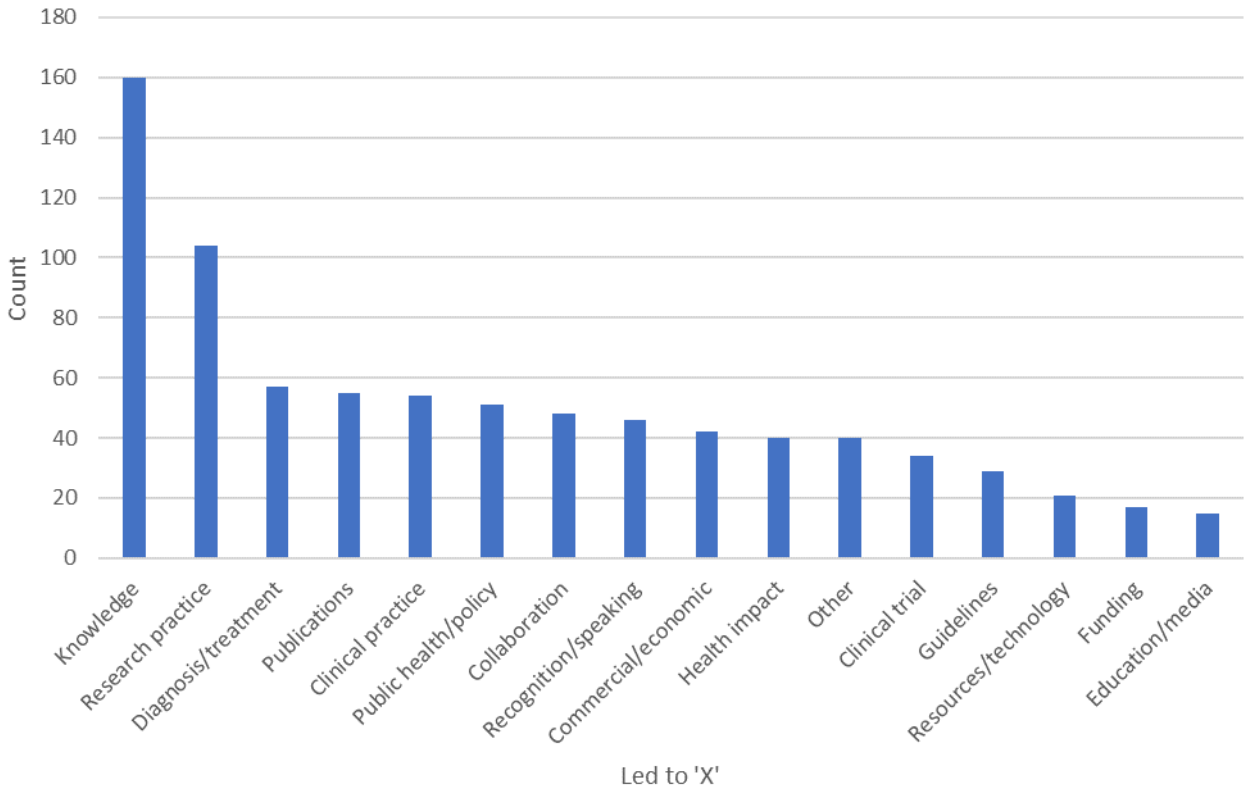
Figure 9. Count of most commonly reported ‘novel’ developments



The word 'novel' occurred within 780 substrings. As shown in Figure 9, the most commonly reported novel developments were therapeutic (62 instances) followed by methods (25 instances) and approaches (24 instances). Other common novel developments included various kinds of 'anti-' (e.g. anti-malarial, antibiotic) and 'bio-' (e.g. bioinformatics, bio-signal).

Within the research and significance text the phrase 'led to' occurred 813 times. The impacts being reported were categorised as shown in Figure 10.[13]

Figure 10. Count of types of impacts reported as being 'led to'



Common impacts reported were on knowledge (160 instances), research practice (104 instances) and diagnosis/treatment (57 instances). Three examples from each category are shown below, with the text modified (with the letters X and Y replacing text) to maintain confidentiality.

Knowledge

- new understanding of the mechanism of action for X enzymes
- 1000s of researchers, practitioners, and policy makers having new knowledge
- two new discoveries: a paradigm shift in analysing X flow … and a unique biomarker

Research practice

- the global uptake of methods that I pioneered
- the development and evaluation of the X intervention
- a global paradigm change in providing evidence for efficient and effective delivery of X care

[13] The categorisation scheme used was developed for the present purpose and has no external status. Although, in any instance, a number of outcomes may have been reported as being 'led to', only the first was used for ease of analysis.

Diagnosis/treatment

- led to the production of a series of analogues, including of X, that have improved biological properties
- led to the progression of new drugs to clinical trials, and provided new knowledge and tools for future drug discovery efforts
- led to new diagnostic biomarkers and targeted therapies

Publications

- >30 publications, with 90% of these published in highly ranked international journals
- 10 high-quality papers
- 13 empirical papers (11 1st author; 1 senior)

Clinical practice

- a change in international clinical practice with specialists now having more objective standards to improve their practice
- profound changes in X care in Australia and internationally
- the adoption of a simplified surgical technique, achieving effective X control, with less side effects

Public health/policy

- a paradigm change in terms of health services and knowledge
- official government approval in several countries/territories
- significant social impact by influencing the policies of governments in many countries

Collaboration

- a collaboration with a research institution in X and a pharmaceutical company
- a collaboration with X, a major industry partner
- collaborations with clinics in UK and Germany

Recognition/speaking

- me being the International X Federation Chairman of the Task Force on Y
- my appointment to the scientific advisory board of X
- 44 conference invitations

Commercial/economic

- a startup company, X, which has received over $Y million in venture capital funding
- an industry collaboration to develop assays
- two patents, licensing of the technology to X

Health impact

- a >50% reduction in the case-fatality in hospitals
- a 45% reduction in falls in people with X disease
- significantly improved well-being and quality of living for the participants

Clinical trial

- a new clinical trial for treating X patients
- an RCT to train hospital staff in X rehabilitation practices
- the current national trial of personalised X care

Guidelines

- adoption of clinical guidelines for X disease

- changing, supporting, and updating X guidelines, policies and practice nationally and internationally
- evidence-based revision of Australian X exposure guidelines

Resources/Technology

- the development of new and improved software tools
- my invention of the synthetic X substrate
- the development of a world-first quality assurance device.

Funding

- >$4m in independent funding to develop a targeted-X manufacturing platform
- a fellowship to Canada to develop analytical tools for the assessment of X
- a successful ARC discovery projects grant

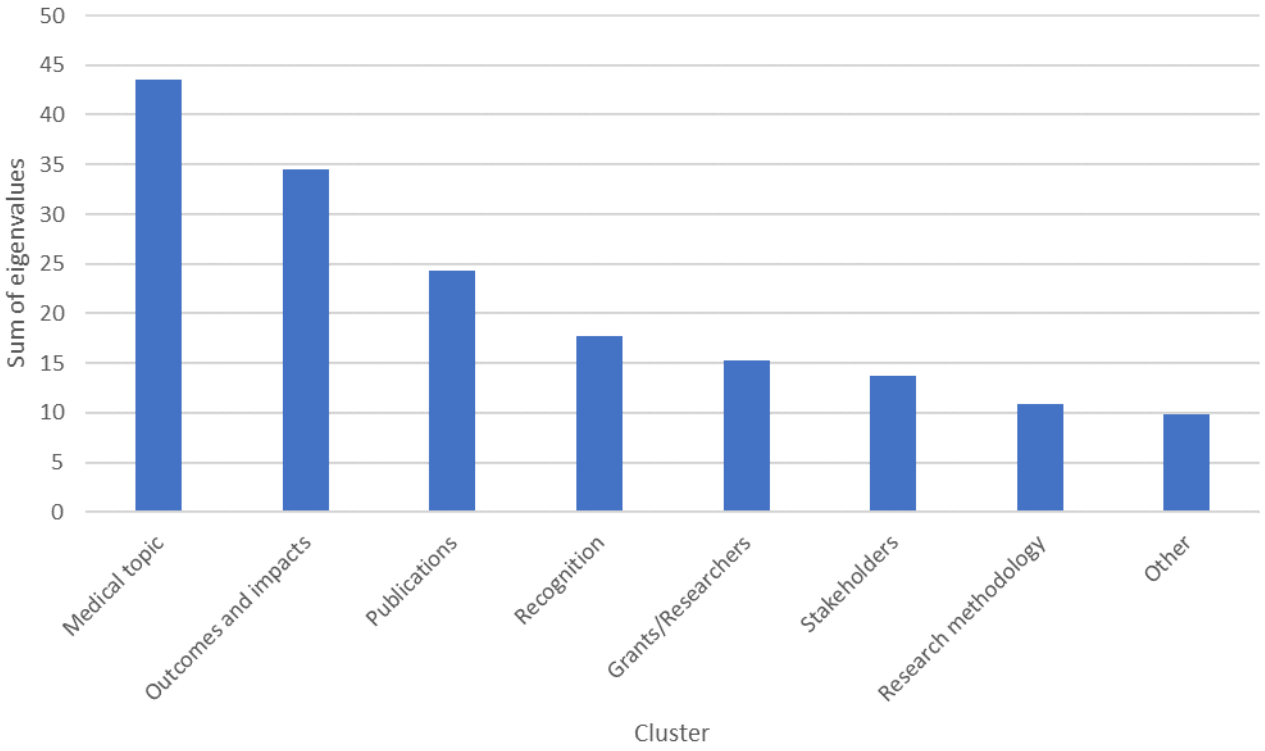
Education/training/Media

- the development of a framework and educational package
- 25 tweets, 3 news articles and cited in 5 countries
- significant media attention which is estimated to have reached 12 million people.

Factor analysis

Another way to understand the content of the impact text is to use factor analysis to identify those groups of words (topics) most commonly found together. Appendix F provides a list of the top 100 topics found within the total impact text. Each topic has been grouped within a broader topic cluster and each topic is also accompanied by an 'eigenvalue', being a measure of how prevalent that topic is within the text.

Figure 11. Topic clusters found in total impact text



As shown in Figure 11, the most prevalent topic of discussion within the impact text is the medical topic that the impact relates to (e.g. 'cancer, breast, prostate, tumour', 'kidney, chronic, disease, renal'). Other topic clusters include:

- Claims about outcomes and impacts (e.g. developed, tools, methods, assessment, software, validated')
- Publications (e.g. 'cited, times, highly, influential, paper')
- Recognition (e.g. 'invited, keynote, speaker, plenary, presentations, talks, conference, international, symposium)
- Grants/researchers (e.g. NHMRC, CIA, grant, project, CIB, APP, GNT, funded, CRE, MRFF, Ideas, ARC')
- Stakeholders (e.g. 'aboriginal, indigenous, communities, Australians, families, children')
- Research methodology (e.g. 'randomised, controlled, trials, RCT').

The value of the combined scores of the Outcomes and impacts (34.5) and Stakeholders (13.7) topics – which together represent impacts beyond academia – is 29% less than the combined scores of the Publications (24.3), Recognition (17.68), Grants/Researchers (15.3) and Research methodology (10.9) – which together represent impacts within academia.

Findings

Applicants are reporting more on knowledge impacts occurring within the research sector than they are on benefits experienced by stakeholders beyond the research sector. This is predictable given that researchers are not usually responsible for delivering benefits to stakeholders beyond the research sector.

A4 – Is the impact text duplicating the publications text?

Background

The Investigator Grants 2022 Guidelines state the following regarding the type of information that applicants should provide about their publications:

> Applicants are required to nominate up to 10 of their best publications from the past 10 years (taking into account any career disruptions). You are to provide separate explanations for each citation (publication) entry. Each explanation should explain why the publication has been selected, including its quality and contribution to science, and your contribution to each. [Highlighting added]

With regard to the type of information to be included about the applicant's impacts, the Guidelines state:

Research impact

Applicants are assessed based on:

1. the significance and reach of their claimed research impact (7%)
2. the contribution of their research program to the research impact (6%)
3. the contribution of the applicant to the research program (7%).

> NHMRC defines the impact of research as the verifiable outcomes that research makes to knowledge, health, the economy and/or society. Impact is the effect of the research after it has been adopted, adapted for use, or used to inform further research. [Highlighting added]

An explanation of the 'contribution to science' of an applicant's publications might look very similar to a statement about an applicant's outcomes on knowledge and since 91% of all applications in the dataset claimed a knowledge impact, most applicants would be expected to include a statement about their knowledge impact within their impact text.

Indeed, and as demonstrated above (refer Table 3), whether or not applicants selected knowledge impact as an impact type they included a comparatively large number of knowledge keywords in their impact text.

Question:

1. Is there duplication of text between the publications and impact sections?

Methodology and results

This question has been examined in three different ways:

1. Through survey questions asking applicants and peer reviewers about the overlap between the impact and publications sections. These were:
    a. Applicant questions 8, 9 and 10
    b. Peer reviewer questions 19 and 20.
2. Examining overall word frequencies across the research impact and publication texts.
3. Examining whether publication dates are being provided in both blocks of text.

Surveys

Both applicants (Q8) and peer reviewers (Q19) were asked to rate the degree to which the information that applicants provided for the three impact components overlapped with the publications section.

As shown in Figure 12:

- 53% of applicants thought that these two texts overlapped either moderately or a lot, while 47% thought that they overlapped a little or not at all
- 62% of peer reviewers thought that these two texts overlapped either moderately or a lot, while 35% thought that they overlapped a little.

Applicants were also asked (Q9) how many of their top 10 publications were used as evidence in their research impact section. As shown in Figure 13, 56% reported listing between 5-10 of their top 10 publications within the impact section, while 42% reported listing 1-4. Only 1% reported not including any.

Figure 12. Applicant and peer reviewer responses to Q8 and Q19 (respectively), on whether the information provided in the publications section overlapped with the information provided within the impact section

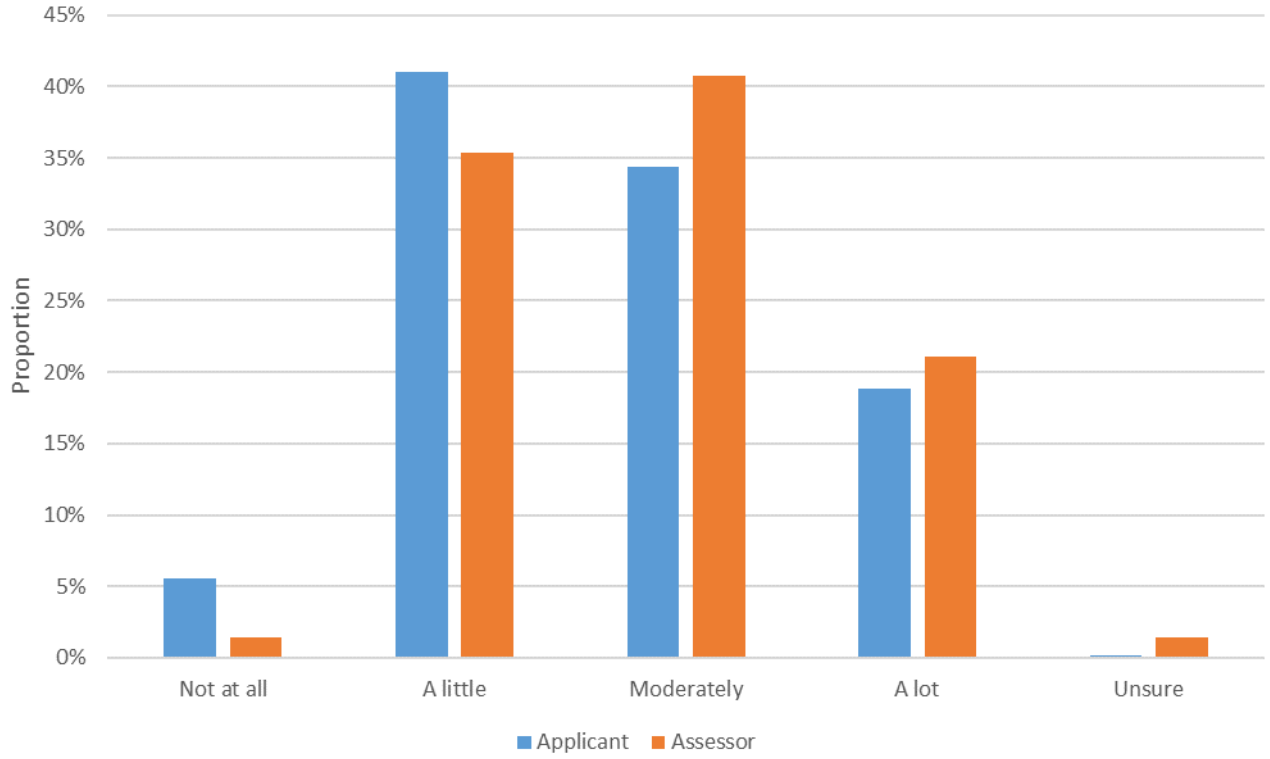
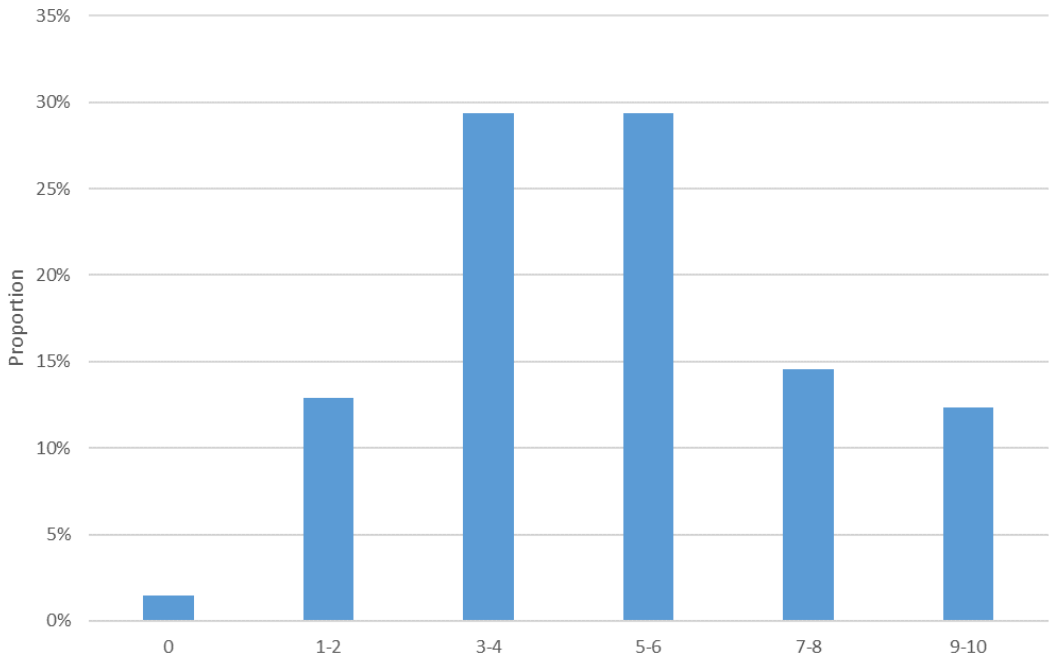


Figure 13. Applicant responses to Q9: "How many of the publications listed in your top 10 publications were used as evidence in your research impact section?"



Both applicants (Q10) and peer reviewers (Q20) were provided with the opportunity to provide additional comments about overlap.

Applicant survey comments

Of the applicant respondents, 35.2% (241 of 684) provided a comment in response to Q10. Responses to this question were varied and respondents took the opportunity to discuss issues with the publication section that were not relevant to the completion of the research impact section. Despite this, some comments were able to provide further context to the responses in questions 8 and 9. This included:

- 10% (24) of respondents discussed track record and a large subset indicated that they found the updated advice to not include any additional track record information limiting.
- Some respondents indicated an understanding that overlap between the publications and research impact sections is to be expected, whereas others indicated that they tried to avoid overlap between these two sections.
- Some respondents explained that, as impact takes a long time to generate, publications used as evidence in the research impact section differed from those used in the publications section as they were outside of the 10-year time limit of the publications section.
- Another subset of respondents (14.5% or 35 of 241) discussed expectations of peer reviewers and that there is conflicting and contradictory advice provided by NHMRC and peer reviewer feedback relating to the use of publications as examples of evidence. These comments suggested that peer reviewers are not provided with appropriate guidance from NHMRC on how to score the publication and impact track record sections and on what is appropriate or allowable in each section.
- An additional 15 respondents suggested that ECRs should be required to only list their top 5 publications, as opposed to top 10.

Peer reviewer survey comments

Of the peer reviewer respondents, 38.8% (109 of 281) provided a comment in response to Q20. Responses discussed a range of issues with the publication section, not all of which were relevant to the research impact section. In total, there were 42 responses (38.5%) which contained comments relating to the impact track record section.

Of those commenting on impact track record assessment, the majority of respondents (68.3% or 28 of 42) commented on themes relating to repetition and overlap. Key comments included that there was overlap between the sections (7.1%), including repeated information (4.8%). However, others reflected that a little overlap was unavoidable (7.1%), that there was varied overlap between sections (4.8%) and that the amount of repetition was dependent on the applicant (7.1%), with the best applicants having minimal overlap (4.8%).

Data analysis

Word frequencies

A word frequency analysis was performed on each of the research impact and publication texts for 6,753 applications from 2019-2022. Collectively, all instances of research impact text contain about

4,100,000 words[14] and about 72,000 distinct words, while all instances of publications text contain about 1,500,000 words and about 65,000 distinct words.[15]

Table 5 was compiled by finding the top 100 words by tf-idf for the publications text, and then removing all words that did not specifically relate to publication in general. This included removing the names of various journals (such as The Lancet). This process left 25 words, whose tf-idf scores are shown for both the Publications text and for the total Impact text. As can be seen, in 10 cases the tf-idf score is greater for the impact text than for the publications text. In other words, even though these words were isolated by rank-ordering the publications text, they are actually more present within the impact text.

Table 5. Word frequencies in publications and impact text

#	Word	tf-idf impact	tf-idf publications	Greatest tf-idf value
1	DOI	0.32	0.91	Publication
2	FWCI	0.62	0.64	Publication
3	JOURNAL	0.25	0.47	Publication
4	PAPER	0.28	0.38	Publication
6	SCOPUS	0.22	0.27	Publication
8	ALTMETRIC	0.16	0.26	Publication
9	CITATION	0.18	0.22	Publication
10	PERCENTILE	0.10	0.22	Publication
11	ARTICLE	0.17	0.22	Publication
15	WEIGHTED	0.12	0.20	Publication
21	GOOGLE	0.18	0.19	Publication
29	AUTHORS	0.16	0.19	Publication
33	RANK	0.05	0.18	Publication
34	SCORE	0.12	0.17	Publication
36	MANUSCRIPT	0.15	0.16	Publication
37	AL[16]	1.22	0.69	Impact
40	CITES	0.73	0.64	Impact
54	PMID	1.24	0.43	Impact
59	CITATIONS	0.39	0.37	Impact
61	TOP	0.40	0.34	Impact
66	CITED	0.41	0.24	Impact
83	AUTHOR	0.41	0.23	Impact

[14] The term 'word' is being used flexibility here to also include initialisms such as FWCI (field weighted citation impact). Plural versions of a word are counted as distinct words in this analysis (e.g. publication and publications, journal and journals). Words exclude frequently occurring parts of speech, as noted in Section 1.

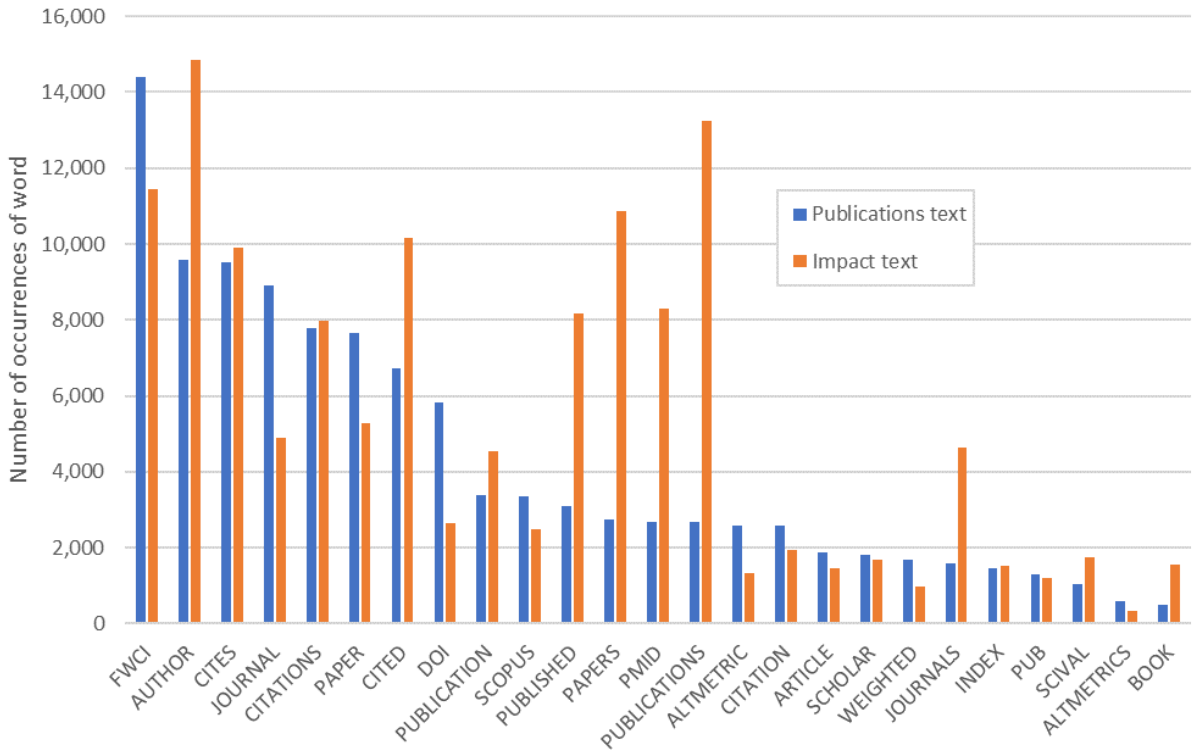
[15] When calculated by WordStat removing commonly used parts of speech. When counted within Excel there are about 6.82 million 'words' in the impact text and 2.47 million in the publications text.

[16] As with the text analysis reported in Table 4, 'AL' comes from 'et al', a frequently used expression when describing a publication.

#	Word	tf-idf impact	tf-idf publications	Greatest tf-idf value
85	CIT	0.35	0.22	Impact
87	PUBLICATION	0.32	0.17	Impact
88	FIELD	0.43	0.17	Impact

Figure 14 shows the number of occurrences of the top 25 most frequent publications-related terms in both the publications and impact text. As can be seen, most terms occur about as frequently in each type of text, with the exception of the words 'author', 'cited', 'published', 'papers, 'PMID', 'publications' and 'journals', which are used as or more frequently in the impact text than in the publications text.

Figure 14. Number of publications-related words in publications and impact text



Publication dates

While the analysis above shows that publication-related language is being used in both the publications and impact text, it does not reveal whether the same specific publications are being referred to in both sets of text. This is difficult to ascertain directly because applicants use a variety of inconsistent approaches when referring to their publications and consequently it is not possible to search, for example, for the same entire citation details in both the publications and impact text.

It is, however, possible to look for a proxy, being the year of a publication. While citation information does not always refer to the publication year (e.g. in some cases a publication is

identified by volume number), the publication year is frequently used as part of the description of a publication.

If applicants are referring to their publications in both the publications and impact text then parallels in the number of times each year is mentioned should be apparent. Conversely, if the years mentioned in the impact text are unrelated to the years mentioned in the publications text then there should be little or no correlation.

Searching on the text strings for each year (e.g. 2020, 2015,) preceded by a bracket (i.e. "(2020") or by a space (i.e. " 2020") to exclude false positives (e.g. "DOI 45782020672") returns the result shown in Figure 15.

Figure 15. Proportion of mentions of years in the impacts and publications texts

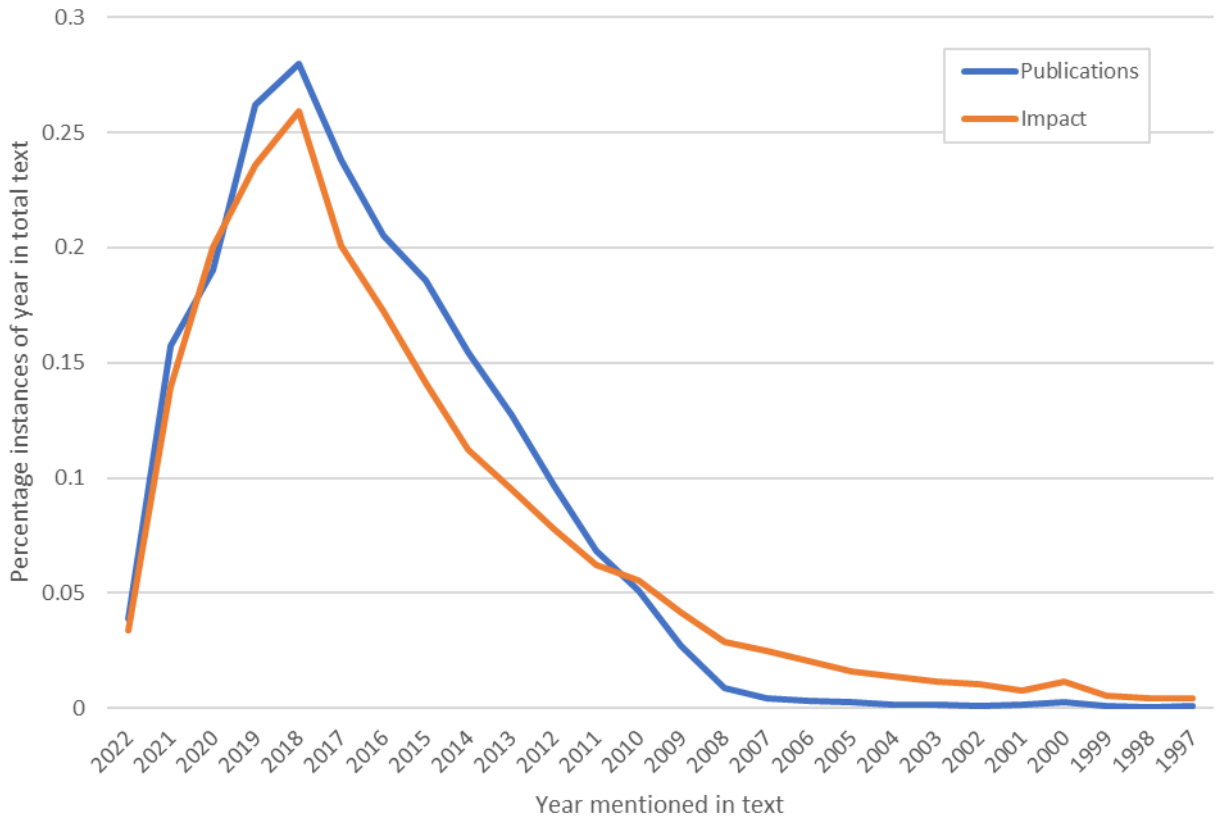
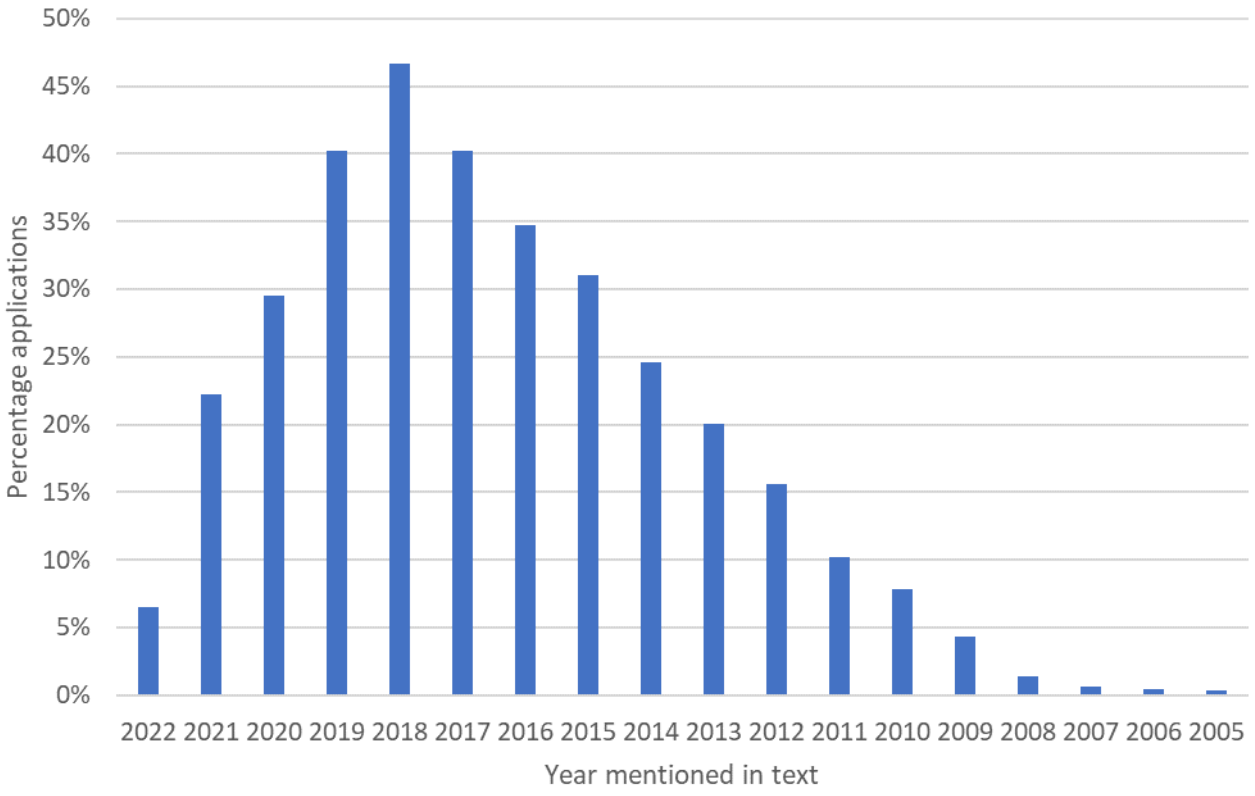


Figure 15 shows the proportion of mentions of a given year in the publications and impact texts (as a proportion of all words in each block of text). It shows that these proportions are very similar. It also shows that the number of mentions of years is behaving as might be expected in theory. That is:

- both lines peak in 2018, a year that applicants to all rounds might reasonably be expected to include no matter how long or short their career had been, or which round they had applied to (noting that applicants would rarely be able to mention years later than the round year)
- the publications line rises linearly to 2018 from about zero in 2008 (a year which no applicant was supposed to include given the 10-year retrospective inclusion period, unless they had a career disruption).

Figure 15 suggests a strong correlation between the mentions of a given year within each text, but it does not show that the same application mentions a given year in both texts. To make sure that this is the case, Figure 16 shows the percentage of applications within which the same year is mentioned in both texts. For the most recent year available for all applications (2018), this figure is almost 45%, dropping to less than 5% for the earliest permissible year (2009).

Figure 16. Percentage of applicants who mention a given year in both the impact and publications text



The pattern shown in Figure 16 is very similar to that shown in Figure 15, further suggesting that the strong overlap between the years mentioned in both texts is due to the same publications being referred to in the impact text.

Findings

The analyses above provide clear evidence of a strong overlap between the publications and impacts text. Applicants report having created this overlap, peer reviewers report seeing it and an analysis of the data independently confirms that it exists. Furthermore, evidence provided previously has shown that it is knowledge impacts that are most frequently being reported within the impact text, thus overlap should be expected since it is also knowledge impacts that are supposed to be reported within the publications text.

B1. How easy was it to assess the impact text?

Background

Peer reviewers can find it difficult to score impact text, just as applicants can find it difficult to draft impact text. The Investigator Grant Guidelines provide category descriptors to assist peer reviewers to provide impact scores.

Questions:

1. How helpful did the peer reviewers find the category descriptors?
2. How easy was it to assess the impact text?

Methodology

A number of questions within the peer reviewer survey related to the ease of assessment of the impact text.

Peer reviewers were asked to rate:

- how helpful they found the category descriptors when assessing the research impact section (Q15).
  - Peer reviewers were also able to provide comments on the category descriptors for the three components of the research impact section (Q16)
- the ease with which they could assess each impact component (Q24)

Peer reviewers were also asked, when assessing the research impact section, whether they read all the text first and then scored the components, whether they read the text of each component and scored it before moving to the next, or some other approach (Q26).

Peer reviewers then had an opportunity to provide their thoughts about the research impact section (Q27).

Results

As shown in Figure 17, regardless of the text component, peer reviewers found the Category Descriptors to be more unhelpful than helpful when assessing the applicant's research impact section. On average, 64% of peer reviewer respondents found the Category Descriptors to be Unhelpful or Very unhelpful, compared with only 16% that found them Helpful or Very helpful.

Of the peer reviewer respondents, 44.5% (125 of 281) provided an additional comment (in response to Q16). The relatively large frequency of responses to this question compared to other free text fields relating to the research impact track record is indicative of the relative difficulty of assessing this section and the interest peer reviewers have for adjusting the current assessment system.

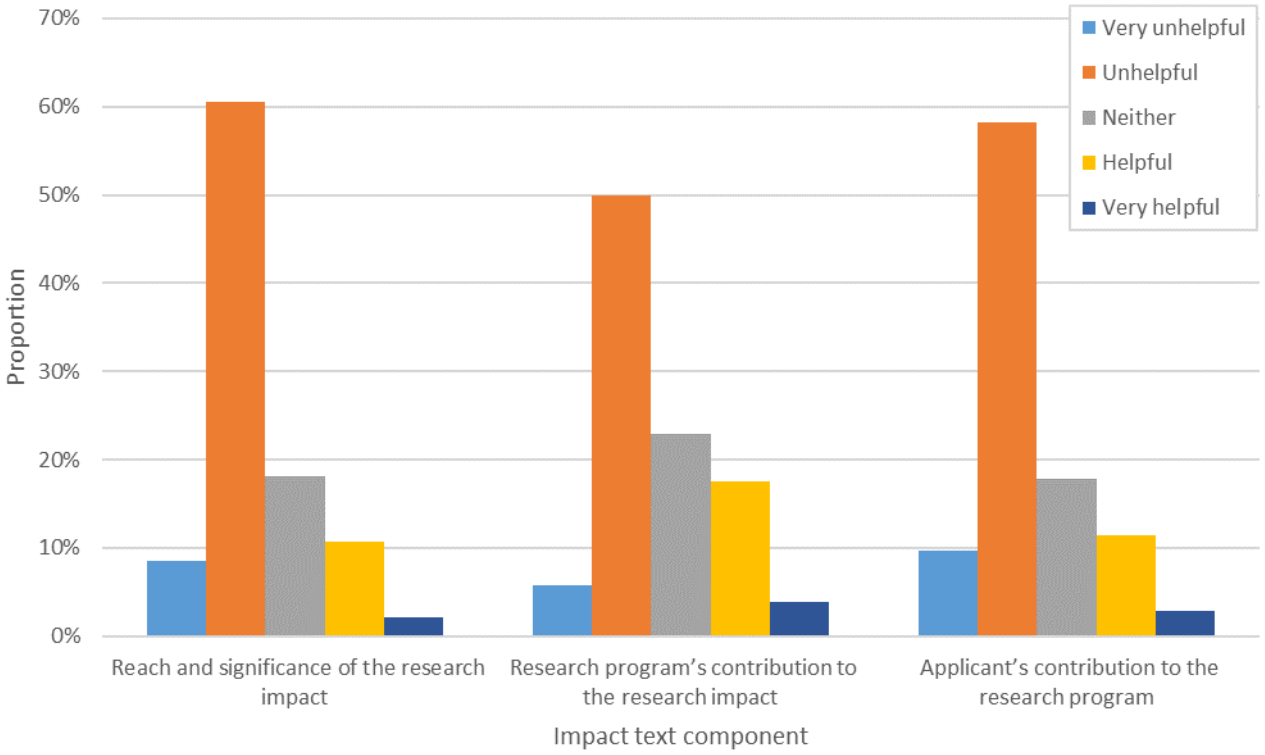
Of these respondents, 36.8% commented on aspects related to scoring, 34.4% on impact sections and 20% on category descriptors.

Key comments discussed:

- That the category descriptors (4%) and scoring (2.4%) are subjective. Comments related to the use of adjectives such as "outstanding" and "excellent" as distinguishing features of categories.

- The perceived disadvantage of some disciplines in the impact sections (4.8%) due to difficulties of scoring across disciplines, low FWCI and types of impacts that can be achieved.
- Perceived biases in scoring which favour researchers who come from large research groups (1.6%).

Figure 17. Peer reviewer responses to Q15: "How helpful were the category descriptors for each component" by impact text section



Respondents provided the following key suggestions for improving the peer review of the impact track record section:

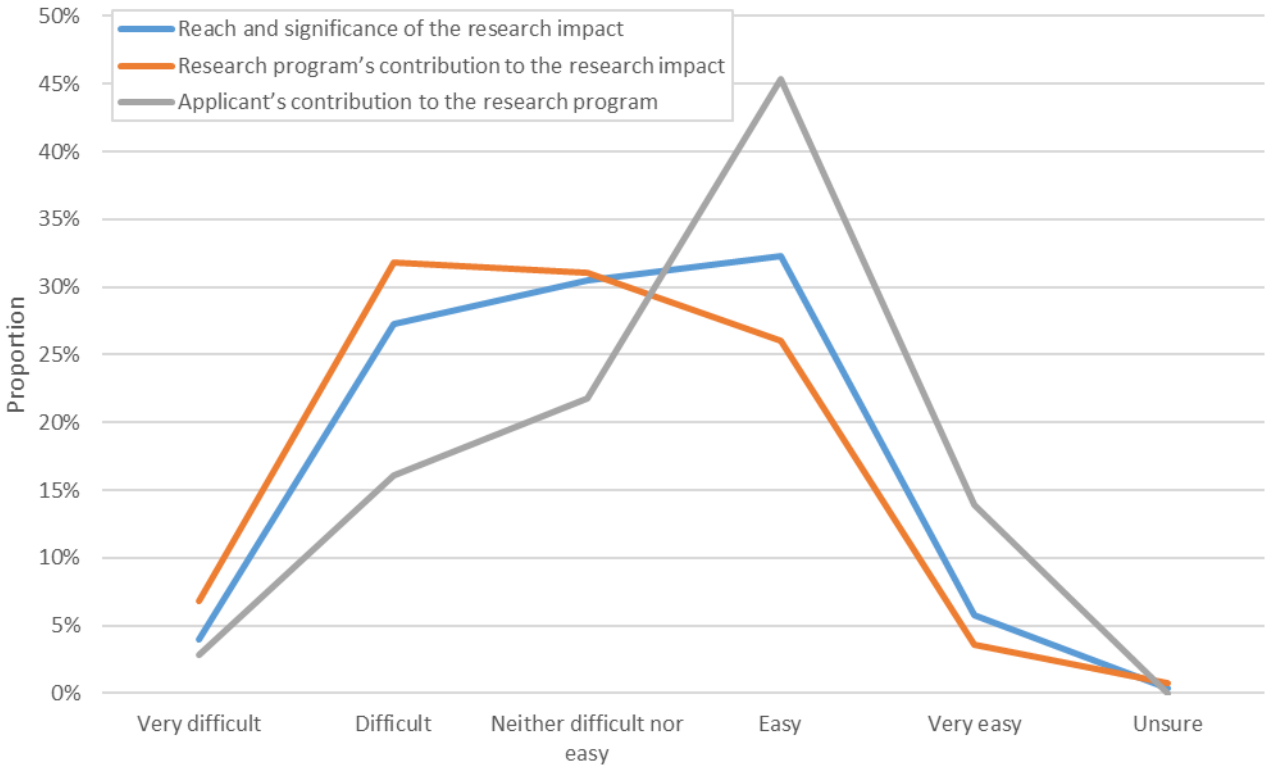
- Changing the number of impact track record subsections (17.6%) by:
  - Merging all sections into a single impact field (13.6%), merging subsections 1 and 2 (1.6%), simplifying sections (1.6%) or removing a section (0.8%). This was felt by some respondents as a possible way to address the concern raised by 4.8% of respondents that the sections were often confused by applicants.
- Increasing the clarity of category descriptors (14.4%), as 8% described the current differences as too subtle, making scoring difficult.
- Provision of example responses at each score and across each applicant level to aid consistency and ease of scoring (8.8%).
- Increasing section clarity (7.2%) by clearer delineation of sections and provision of clear definitions of sections.
- Increasing the range for scoring (4.8%) either by expanding the 1-7 scoring scale or introduction of half marks.

Other minor suggestions included:

- The use of panel review (4.8%) or adjusting scores after review of shared peer reviewer comments (0.8%).
- Provision of examples of responses to applicants (2.4%) to improve application quality.

As shown in Figure 18, peer reviewers found assessing the research impact components to be of varying degrees of difficulty. Peer reviewers found the applicant's contribution to the research program the easiest component to assess, while assessment of the other two components was reported to be both easy and difficult.

Figure 18. Peer reviewer responses to Q24: "How easy was it to assess the following components?"



That said, the majority of peer reviewers (66%), when assessing the research impact section, read all the text first and then scored the components, while only 34% read the text of each component and scored it before moving to the next (Q26).

Of the respondents, 34.9% (98 of 281) provided an additional comment on the research impact section (Q27). Responses covered a range of broad themes including section structure (31.6%), scoring (22.4%), equity (16.3%), execution (15.3%), guidance (14.3%) and overlap/repetition (14.3%).

These comments discussed:

- that this section was hard to assess and difficult to score (7.1%), with 3.1% remarking that this was the hardest section to assess

- that applicants were often confused as to which impact type to choose and what information to include in each of the sections (7.1%), leading to applicants placing content in the wrong sections (3.1%)
- that there was repetition or overlap between impact types, sections and with the publication section (14.3%)
- that the current approach favours style over substance, with score relating to writing skills not scientific quality or rewarding those who are well coached (5.1%)
- a feeling that there was inequity between career levels with established researchers advantaged and EL applicants disadvantaged by the current system (4.1%)
- a belief that there was inequity between different disciplines or fields (4.1%)
- that applicants from large research groups were advantaged (3.1%) due to increased opportunities for impact by participation in large programs.

Some respondents offered suggestions for improvements across the three free-text field components.

Key suggestions included:

- merging of the impact sections into one section (18.4%) or changing the order of the three impact sections (3.1%)
- increasing the space/word count for responses (4.1%)
- provision of more direction for applicants (14.3%). This included the provision of clearer examples including those across varied fields, direction regarding section content, and impact type descriptions.

Findings

Peer reviewers generally found the Category Descriptors to be unhelpful, they found the applicant's contribution to the research program the easiest component to assess, while they found assessment of the other two components to be both easy and difficult.

B2. Do applicant characteristics influence scoring of impact?

Background

It is possible that applicant characteristics might be associated with differences in scoring because the RITRA framework is structured in ways that suit the needs of some types of applicants better than those of others. These characteristics might include gender (e.g. Tamblyn et al 2018), ethnicity (e.g. Taffe and Gilpin, 2021[17]), age or level of seniority, career disruption or research area.

Question: Does any evidence exist to suggest that particular categories of applicants are advantaged or disadvantaged by the framework?

[17] Taffe MA and Gilpin NW. Equity, Diversity and Inclusion: Racial inequity in grant funding from the US National Institutes of Health. eLife. 2021 10

Methodology

Applicants to the Investigator Grant scheme provide textual, categorical and numerical information within their applications. The textual information relates to each applicant's publications, leadership track record, research impact and project proposal. The categorical and numerical information (the profile information) includes:

1. Application year (AY), coded in the dataset as 2019=1 to 2022=4
2. Broad Research Area (BRA) coded as four separate variables, noting that applicants can choose only one BRA:
    a. Basic Science (BS) (1,0)
    b. Clinical Medicine and Science (CM) (1,0)
    c. Health Services (HS) (1,0)
    d. Public Health (PH) (1,0)
3. Indigenous research (In), coded Indigenous=1, Other=0
4. Gender (Gn), coded Female=1, Male=0. Other values were removed from the analysis due to low numbers.
5. Years post PhD (YP), containing the number of years between the PhD date and the application year.
6. Career disruption (CD), containing the number of days of career disruption.
7. Leadership level (LL), coded linearly (1-5: Emerging Leadership level 1 = 1 to Leadership level 3 = 5)
8. Impact type, coded as four separate variables, noting that applicants can choose more than one impact type:
    a. Knowledge (Kn) (1,0)
    b. Health (He) (1,0)
    c. Social (So) (1,0)
    d. Economic (Ec) (1,0).
9. Impact type total (IT), the summation of the values for the four Impact types.

Peer reviewer scores (on a scale of 1-7) are also present in the dataset for publications, leadership, the three research impact components and knowledge gain.

The relationship between applicant characteristics and scoring may be investigated, among other possible ways, by using correlations and regressions to indicate the extent to which other information besides the impact text is predictive of the total impact score.

Results

Correlations

Before investigating correlations between profile information and impact scoring, it is first possible to investigate the level of independence of the profile variables and determine whether knowing

the value of one profile variable allows us to predict the value of another profile variable. To the degree that two profile variables correlate then they may provide different versions of the same underlying information.

Background information on correlations is provided in Appendix G.

A matrix showing cross-correlations among the profile variables is provided at Table 6. Negative correlations (indicating that as one variable increases in value the other tends to decrease) are indicated in shades of red, while positive correlations are indicated in shades of blue. Cross-correlations among the BRA (which are mutually exclusive values and thus must negatively correlate) were removed, as were correlations between the impact types and impact type total. The only correlations included within the matrix are those where $p<0.01$.[18]

The most notable feature of the correlations shown in Table 6 is that they are almost all small, despite being statistically significant. This means that the profile variables are mostly independent of, or have little impact on, each other.

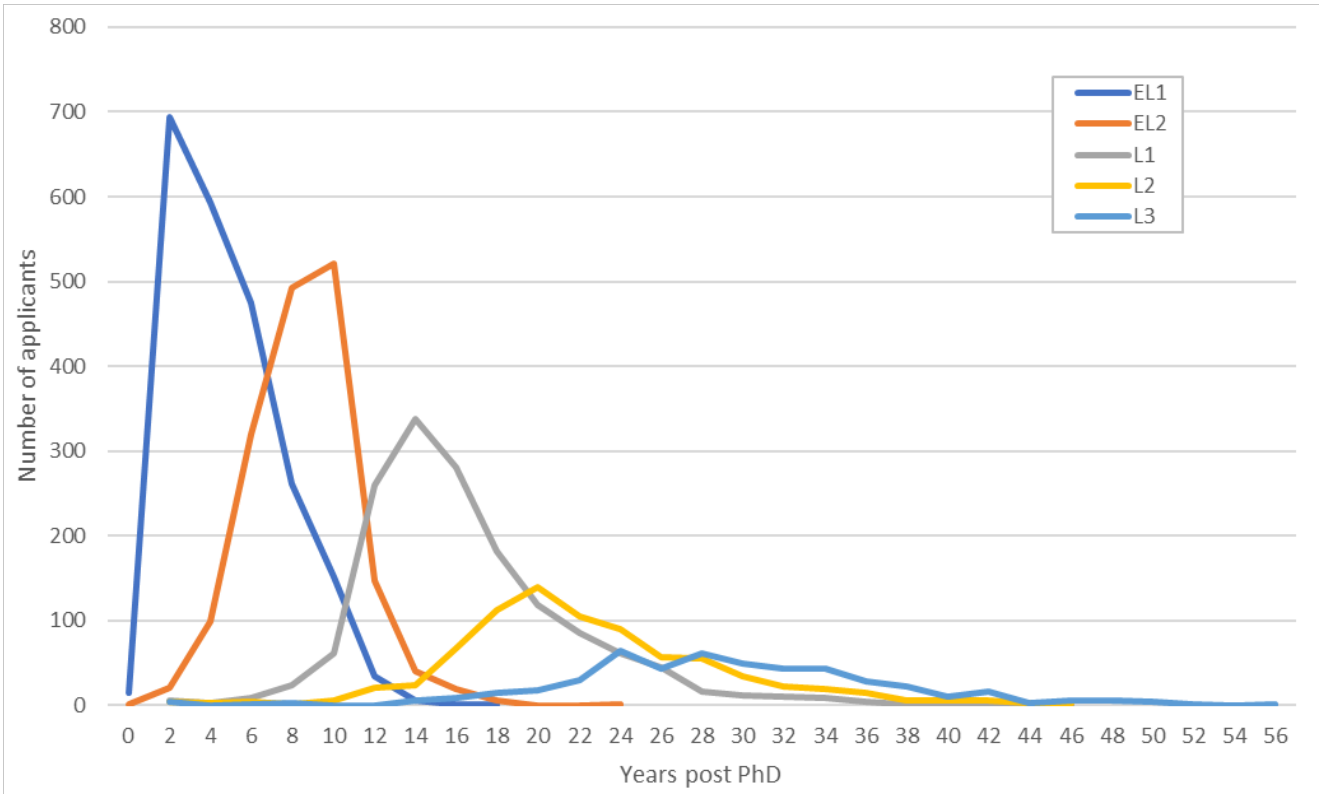
Only one correlation is large: a positive correlation (r=0.862) between years post PhD and leadership level (coloured blue). This relationship, graphically depicted in Figure 19 is predictable given that it takes time for researchers to gain seniority within their profession. It also relates to the eligibility and statements of expectations for the scheme.

Table 6. Correlation matrix for profile variables

	1. AY	2.BS	3.CM	4.HS	5.PH	6.In	7.Gn	8.YP	9.CD	10.LL	11.Kn	12.He	13.So	14.Ec
1. AY	1													
2. BS		1												
3. CM			1											
4. HS				1										
5. PH					1									
6. In		0.138			0.149	1								
7. Gn		0.151		0.148	0.134		1							
8. YP							0.162	1						
9. CD		0.118					0.348		1					
10. LL							0.197	0.862		1				
11. Kn		0.174		0.109	0.128						1			
12. He		0.402	0.228	0.168						0.171	0.288	1		
13. So		0.194		0.124	0.259	0.122					0.167	0.1	1	
14. Ec										0.105		0.119	0.107	1
15. IT		0.299	0.14	0.136	0.1					0.151				

[18] This information is also provided in Appendix H including negative signs for negative correlations, which were omitted from this table due to limitations of space.

Figure 19. Number of applicants by years post PhD by leadership level



A medium-sized negative correlation (r=-0.402) exists between BRA - Basic Science Research and impact type – health (coloured red). This indicates that applicants who chose Basic Science Research as their BRA were less likely to choose health as their impact type. This correlation has some parallels with the small-medium negative correlation (r=-0.288) between the impact types of knowledge and health. Basically, it indicates that the BRAs and impact types are to some degree measuring the same thing or are otherwise aligned. This is predictable since the primary impact to be expected from a Basic Science Research activity is on knowledge.

Another medium-sized correlation (r=0.348) exists between gender and career disruption, indicating that applicants selecting gender=female were more likely to have reported more days of career disruption. This, too, is predictable. As noted by Barnett et al (2022), "A common reason for career disruption is caring for children, and hence career disruption is an issue often keenly felt by female researchers."

Table 7 provides correlations between the profile information variables and the peer reviewer scores. The only medium-sized correlations are positive correlations between years post PhD/leadership level and the leadership and total impact scores (r=0.313, 0.366, 0.302 and 0.356). Such correlations are predictable given that the longer a researcher has been working the more opportunity they have had to engage in leadership activities and for their work to have led to impacts.

Table 7. **Correlation matrix for profile information variables against peer reviewer scores**

	Publications	Leadership	Knowledge gain	Total impact
1. Application year	0.125	0.134		0.152
2. Basic Science			0.118	
3. Clinical Medicine and Science				
4. Health Services			-0.123	
5. Public Health				
6. Indigenous research	-0.137		-0.124	-0.112
7. Gender	-0.135		-0.105	-0.104
8. Years post PhD	0.171	0.313	0.231	0.302
9. Career disruption				
10. Leadership level	0.244	0.366	0.268	0.356
11. Knowledge				
12. Health		0.115		
13. Social				
14. Economic				
15. Impact type total				

Small negative correlations are apparent between gender and indigenous research and the scores for publications, knowledge gain and total impact.

Table 8 shows correlations between the peer reviewer scores. It shows that all of these correlations are large and some are very large, notably:

- research program contribution and reach and significance (r=0.868)
- research program contribution and applicant contribution (r=0.841).

High correlations between the peer reviewer scores are to be expected for at least two reasons. The first is that most (65.6%)[19] peer reviewers read all the text first and then score all the components together. Consequently, they have been influenced by all of the text when scoring any given component.

The second is that all of the blocks of text were written by the same applicant about various elements of the same research career, and thus are likely to have a variety of strengths in common.

Table 8. **Correlation matrix for peer reviewer scores by component**

	Pb	Rs	RP	AC	Ls	KG
Publications (Pb)	1	0.703	0.704	0.713	0.675	0.640
Reach and Significance (RS)		1	0.868	0.789	0.706	0.683
Research program contribution (RP)			1	0.841	0.723	0.697
Applicant contribution (AC)				1	0.767	0.686
Leadership (Ls)					1	0.647
Knowledge gain (KG)						1

Regression on inputs only

Linear regression is a statistical technique used to analyse the direct association between a dependent variable and one or more independent variables. A linear regression tests the changes in the mean (average) of the dependent variable based on changes to the values of the independent variables and it demonstrates which of the independent variables are the best predictors of the value of the dependent variable.

[19] Refer peer reviewer survey Q26

Because independent variables may themselves be correlated (as shown above), by placing them all together in one regression model it is possible to see how much variance in the dependent variable each independent variable accounts for uniquely.

A linear regression was undertaken using the profile information as the independent variables and total impact score as the dependent variable (produced by adding the three unweighted component scores provided for each applicant, providing a maximum possible score of 21).

The regression results (regression coefficients (β values) and their associated significance (p) values) are shown in Table 9. Those independent variables whose p value was >0.01 have been shaded.

The larger a regression coefficient, the stronger the effect that that independent variable has on the value of the dependent variable – in this case, total impact score. However, the size of a regression coefficient is not to be judged in the same way as the size of a correlation coefficient because the variable with the highest regression coefficient is not necessarily the one with the strongest effect (since the units of each variable may be different).

Moreover, all of the regression coefficients are used to create a formula to predict the value of the dependent variable. The predictive value of this formula also determines how important each regression coefficient is.

Table 9. Application data predicting total impact score – linear regression output

Model	β	p	sig
1. Application year	0.233	<0.001	Y
2a. Basic Science Research	0.339	<0.001	Y
2b. Clinical Medicine and Science Research	0.319	<0.001	Y
2c. Health Services Research	[-0.339][20]	<0.001	Y
2d. Public Health Research	0.365	<0.001	Y
3. Indigenous research	-1.063	<0.001	Y
4. Gender	-0.132	<0.008	Y
5. Years post PhD	-0.004	0.413	
6. Career disruption days	0.000	<0.008	Y
7. Leadership level	0.553	<0.001	Y
8a. Knowledge	-.176	0.048	
8c. Social	-0.334	<0.002	Y
8d. Economic	-0.119	0.261	
9. Impact type total	0.150	<0.005	Y

As shown in Table 9, the independent variables that are statistically significant predictors of total impact score (at a p value of <0.01) are Application year, the BRAs, Indigenous research, Gender, Years post PhD, Career disruption days, Leadership level, Impact type – Social, and Impact type – Total.[21] However, the regression model itself accounts for only 16% of the variance of the dependent variable (r^2=0.159) meaning that the model is a weak predictor of total impact score.

[20] Because the BRAs are mutually exclusive they cannot all be placed in the model at the same time. The value for Health Services Research was calculated within a different regression model where Basic Science Research was excluded. Its inclusion here is indicated with square brackets to show that it is not formally part of this model.

[21] Impact type – Health was removed from the model as it exhibited collinearity (i.e. too strong an overlap) with one or more other variables

Consequently, even variables with statistically significant β values have only limited predictive value.

These results suggest that peer reviewers are not being biased by the profile information to any significant extent.

Regression on inputs and outputs

One reason the predictive power of the model above is low is that it omits the key source of information that peer reviewers are supposed to use when they make their assessment decisions: the impact text.

This text is not able to be placed in the model as regression requires numerical inputs. It can also not be added to the model numerically represented by the total impact score as this is what the regression model is intended to predict. However, since the peer reviewer scores for the other blocks of text in the application (i.e. publications, leadership and the project proposal) correlate strongly with the total impact score they could be used as proxies for it. Running the regression model with these peer reviewer scores included leads to the results shown in Table 10. Research impact type total was removed from the model as it exhibited collinearity with one or more other variables.

The R^2 value for this new model is substantially higher (R^2 =0.744) than that of the previous model meaning that it now accounts for almost 75% of the variance of total impact score.

Table 10. Linear regression output

Model	β	p	sig
1. Application year	0.066	<0.001	Y
2a. Basic Science Research	-0.009	0.849	
2b. Clinical Medicine and Science Research	-0.065	0.150	
2c. Public Health Research	-0.050	0.312	
3. Indigenous research	-0.070	0.301	
4. Gender	0.012	0.674	
5. Years post PhD	0.008	<0.005	Y
6. Career disruption days	<0.001	0.403	
7. Leadership level	0.064	<0.002	Y
8a. Knowledge	-0.044	0.340	
8b. Health	0.040	0.172	
8c. Social	-0.048	0.345	
8d. Economic	0.140	<0.003	Y
9. Publications	0.826	<0.001	Y
10. Leadership	0.980	<0.001	Y
11. Knowledge gain	0.888	<0.001	Y

Within this revised regression model, the profile variables Application year, Years post PhD, Leadership level and Impact type – Economic, are the only significant predictors of total impact score.

Application year is a predictor because impact scores have been increasing over time (refer Appendix C, Figure 8) While it is not clear why this is occurring, it is not a sign of peer reviewer bias.

That Impact type – Economic is a significant predictor within the model and could indicate that peer reviewers are biased in favour of this type of impact. It could also, however, indicate that this type of impact is considered more valuable by peer reviewers because of its comparative rarity.

That the Years post PhD/Leadership level variables are significant predictors of total impact score suggests that changes may need to be made to the Category Descriptors for the Investigator Grant scheme. This issue is explored further in the next section.

Findings

There does not appear to be a particular applicant characteristic that is advantaged or disadvantaged by the RITRA framework, however the Category Descriptors may be causing issues with scoring by Leadership level.

B3. What factors affect impact scoring?

Background

Leadership level

Results provided in the section above indicate that leadership level is a predictor of total impact score. One probable reason for this is because of the extended lengths of time that usually occur between research and impacts.

Hanney et al (2015) reported research indicating an average time for biomedical and health research to lead to products, policy and practice of almost 40 years. This figure exceeded a previous figure of 17 years reported by Morris, Wooding and Grant (2011). A meta-analysis of NHMRC's own impact case studies found an average duration of 28 years across the first 36 impact case studies published.

Because of this time lag between research and impacts, the number and types of impacts that an applicant is able to report is likely to be highly dependent upon the length of their preceding research career, which – as shown previously – highly correlates with their leadership level.

If applicants at all career stages had to compete with each other using the same scoring system then those from higher leadership levels would be likely to receive greater impact scores. Murray et al 2016 have argued that ECRs require their own scoring system and, as discussed above, NHMRC uses different scoring between the EL and L levels.

It may be that these different scoring systems are not exerting a sufficient remediating effect. It may also be, however, that this problem has been largely addressed already. This is because the information captured for the R2O and guidance provided in the SofE both changed in 2021.

The SofE was more explicit about years post-PhD and academic appointments for each Investigator Grant level (e.g. "it is expected that L3 Investigator Grant recipients will typically be more than 20 years post-PhD (or equivalent) and appointable at Academic Level E").

For R2O, prior to 2021, applicants provided peer reviewers with details of any relative to opportunity considerations and career disruptions that they had experienced, if relevant. The information provided by applicants wasn't structured and typically focused on the circumstance and impact of the situation. It didn't provide peer reviewers with an outline of the applicant's research opportunities or career stage. Commencing in 2021, however, all applicants were required to include:

1. their career stage based on time period since completion of their PhD or equivalent
2. a structured overview of their research career over the 10-year period up to the closing date of the scheme
3. a career context summary outlining their career circumstances, opportunities for research and the associated impact on their research productivity
4. details of any career disruptions and their impact.

Provision of this information is intended to affect how peer reviewers consider – among other things – how much opportunity applicants have had to generate impacts. Consequently, if this change produces its desired effects then peer reviewers should be considering not so much the generation of impacts, but rather the generation of impacts per unit of research-active time.

Question: How has SofE and R2O implementation affected impact scoring by leadership level?

Impact type

Applicants are allowed to select from 1 to 4 research impact types. The 6,755 applicants in the dataset collectively selected 9,873 impact types. That is, on average each applicant selected about 1.5 impact types.

A relationship between the number of impact types selected and the applicant's leadership level might be expected as it would (presumably) take longer for a researcher to generate a greater number of impacts.

It might also be expected that – all things being equal – the greater the number of impact types selected the greater the associated impact score would be, because it is more difficult to produce impacts across multiple domains.

However, an increased number of claimed impacts might cause difficulties for peer reviewers. Peer reviewers might, for example, find it more difficult to score applications with a greater number of claimed research impacts because they are not sure how to combine scores from a collection of different types of impact. Peer reviewers might also find it more difficult to assess impact types that they are less familiar with (e.g. economic impacts).

Notably, this question might have been answered in the section above, however Impact type – Total had to be removed from the regression model because of collinearity.

Questions:

1. How does leadership level relate to impact type?
2. Do peer reviewers score different impact types differently?
3. Are impact assessment scores affected by the number of impact types claimed?

Methodology

Data analysis was the primary method used to answer the questions above. However, there were also two relevant questions within the peer reviewer survey.

1. Peer reviewers were asked to rate the extent of their agreement with the statement "I was able to deliver consistent assessment of applications across the four impact types" (Q21).
2. Peer reviewers could also provide comments relating to their ability to deliver consistent assessments or provide suggestions for improvement (Q22).

Results

Leadership level

Figure 20 shows the proportion Investigator Grant applicants receiving each total impact score (rounded up to the nearest whole number) by leadership level, but only for 2019 and 2020 applications. It shows five normal (bell-shaped) distributions whose modal (most common) impact scores progressively increase from EL1 to L3, with the exception that the mode for EL2 exceeds that for L1.

Figure 21 provides the same information using 2021 and 2022 data – that is, data collected after the SofE and R2O changes. It shows that the pattern for the 2021 and 2022 data is similar, but the modal scores sit closer together, indicating that peer reviewers are not making such strong distinctions between the leadership levels.

To investigate this apparent change in variability, Figure 22 shows the total number of applicants, grouped by application date (i.e. 2019-2020 versus 2020-2021) receiving each total impact score (rounded up to the nearest whole number). The 2021-2022 scores were adjusted (by subtracting 0.4 from each score) to bring the means of the two groups into alignment (as the mean of the 2019-2020 group was 15.1 and the mean of the 2021-2022 group was 15.5).

Figure 20. Proportion of 2019 and 2020 applicants receiving each total impact score by leadership level

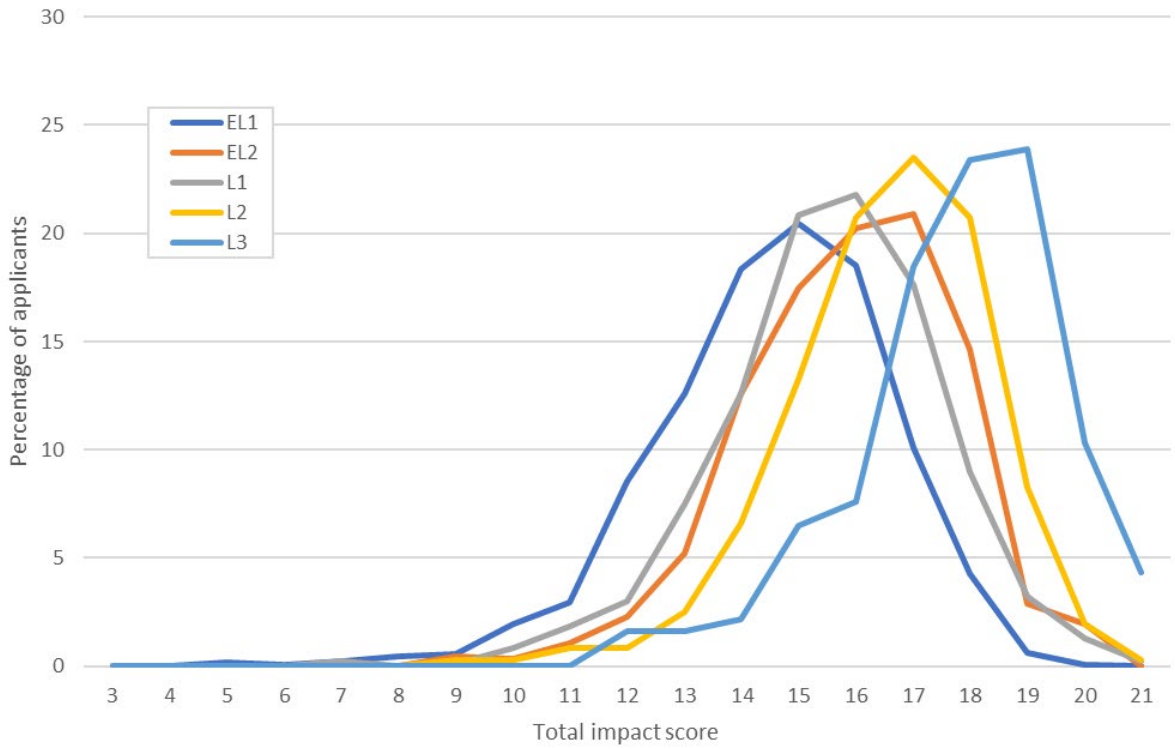


Figure 21. Percentage of 2021 and 2022 applicants receiving each total impact score by leadership level

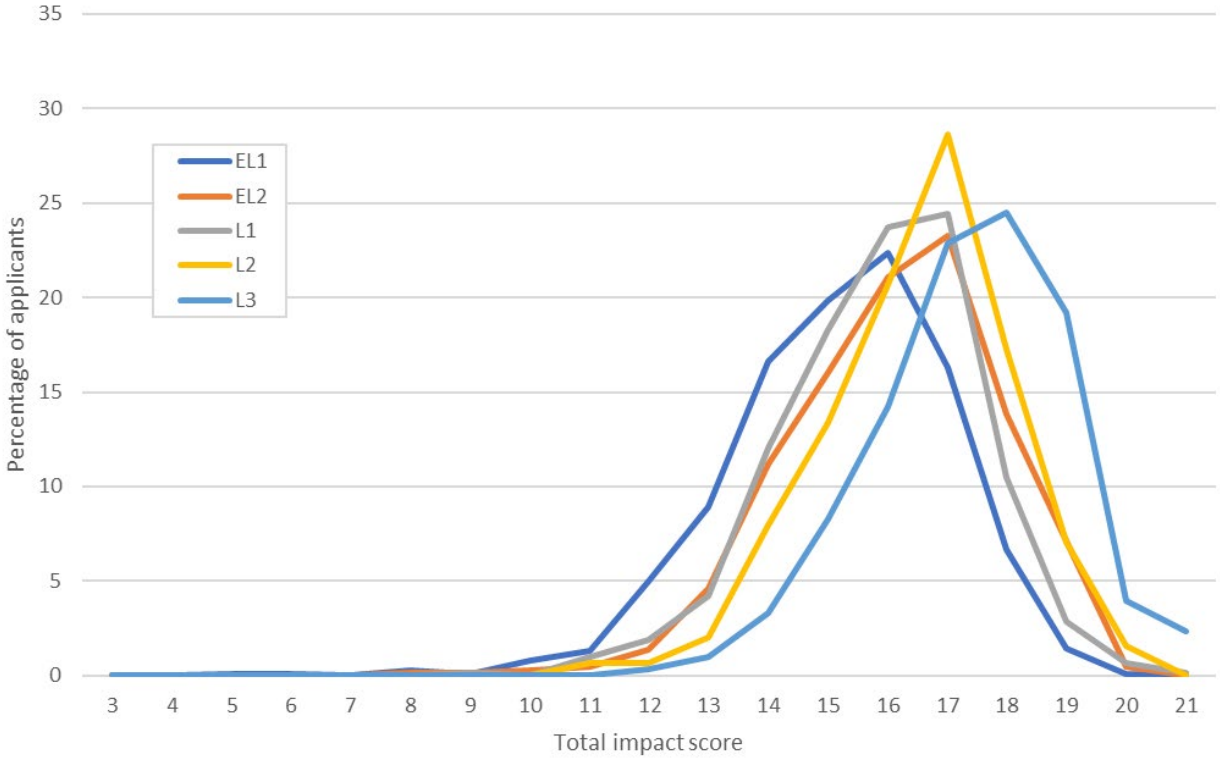
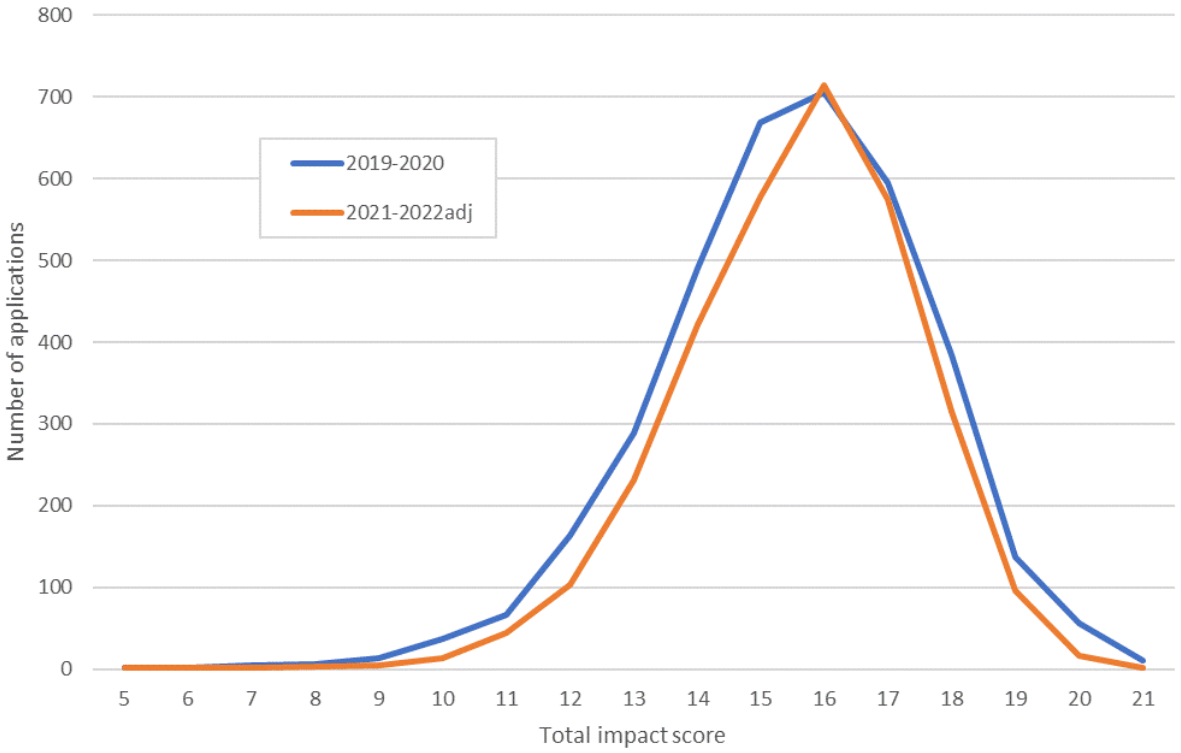


Figure 22. Number of applications receiving each total impact score, 2019 and 2020 versus 2021 and 2022 (adjusted to align means)



As shown in Figure 22, the 2019 and 2020 curve encompasses the 2021 and 2022 curve. That is, the variance of the 2019 and 2020 group is consistently larger than that of the 2020 and 2021 group. As shown in Table 11, an F-test (a statistical method used to test if the variances of two populations are equal) performed on the variances of these two groups was significant at $p<0.01$, indicating that the variances of the two groups are unlikely to be different by chance alone.

This suggests that changes to SofE and R2O have led to a reduction in the disparity between scores at different leadership levels and/or that over time, the peer reviewers gain a better understanding of the requirements and scoring criteria.

Table 11. F test comparing variances of 2019-2020 and 2021-2022 total impact scores

	Mean	Variances	Observations	F	P
2019-2020	15.08	4.391	3,633	1.29	P<0.001
2021-2022	15.54	3.395	3,122		

Impact type

Data analysis

Figure 23 shows the percentage of applicants who selected each impact type by their leadership level. It shows a modest decrease (as leadership level increases) in the number of applicants who selected knowledge, and progressive increases in the number of applicants who selected health

and economic. It also shows that the number of applicants who selected social did not change across leadership levels.

Figure 23. **Proportion of applicants who selected each impact type, by leadership level**

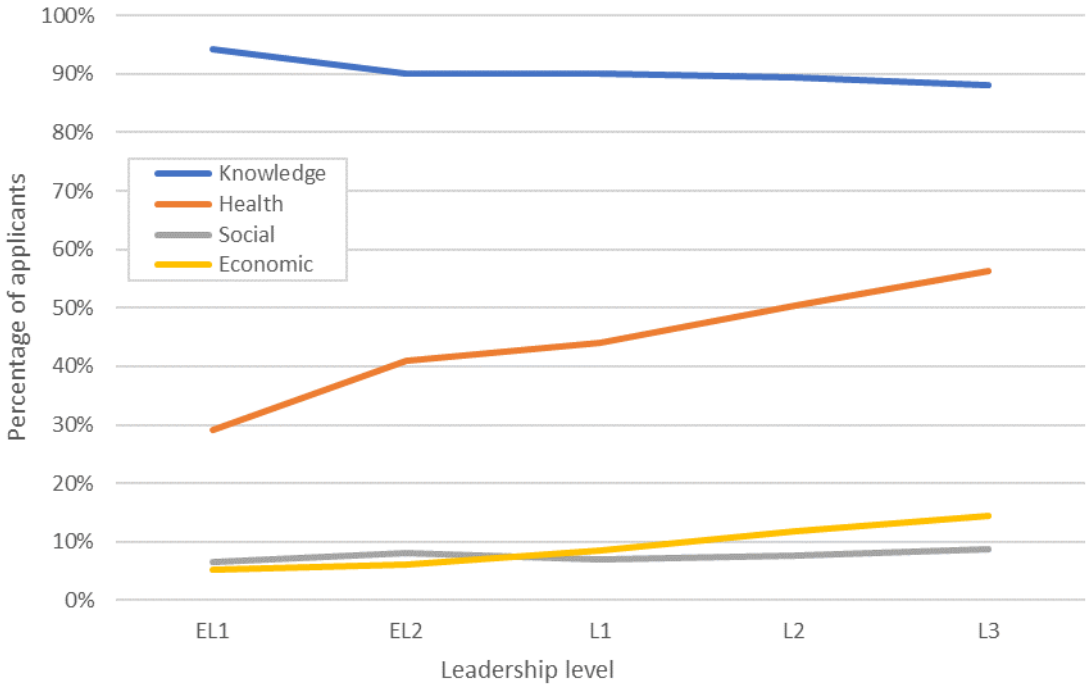
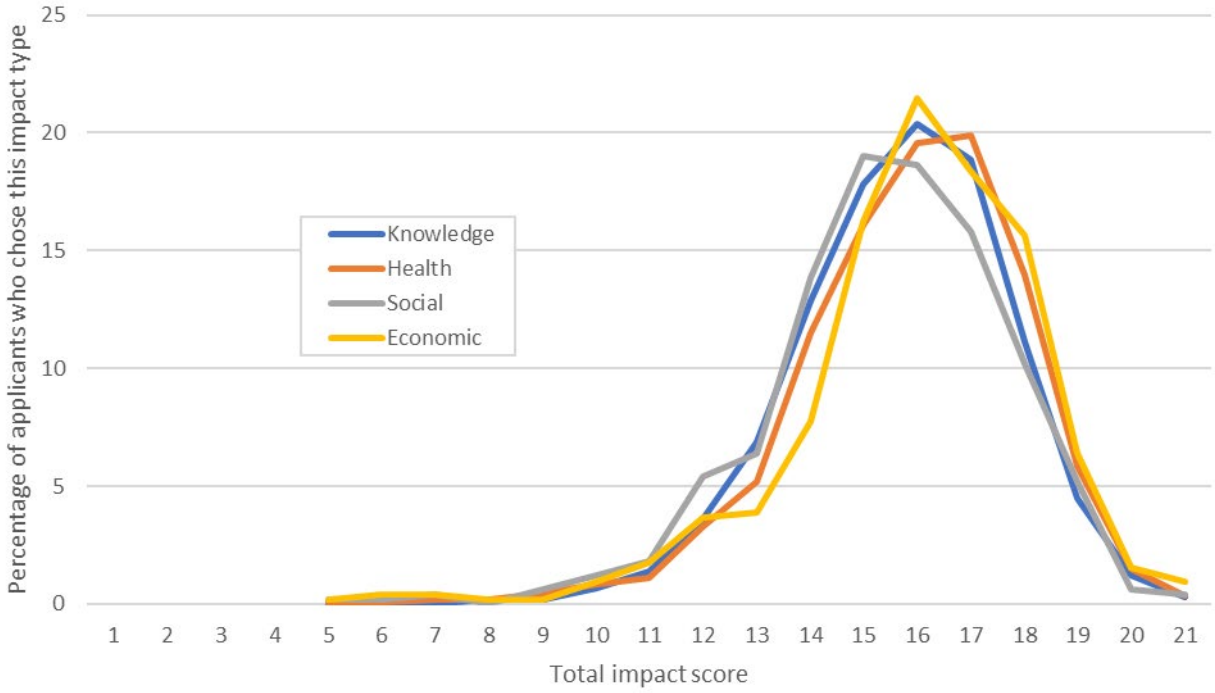


Figure 24 shows the proportion of applicants receiving each total impact score by impact type. It shows that the distributions of these scores are very similar. However, since impact type captures any applicant who selected that particular impact type, regardless of what other types they may have selected, and since 91% of applicants selected the knowledge impact type, there will necessarily be a strong overlap between the other impact types and the knowledge impact type.

The means and the standard deviations of total impact score are similar across the number of impact types selected (means between 14 and 15, standard deviations between 2 and 3), however because of the very small numbers of applications claiming 3 or 4 impacts (358 and 70, respectively) compared with the much higher numbers of applications claiming 1 or 2 impacts (4,131 and 2,194, respectively) it is difficult to determine whether any statistically significant differences exist.

Figure 24. Proportion of applicants at each total impact score level by impact type



Survey

As shown in Figure 25, 55.7% of peer reviewers agreed or strongly agreed with the statement "I was able to deliver consistent assessment of applications across the 4 impact types" while only 13.9% strongly disagreed or disagreed, 26.1% were neutral and 4.3% were unsure.

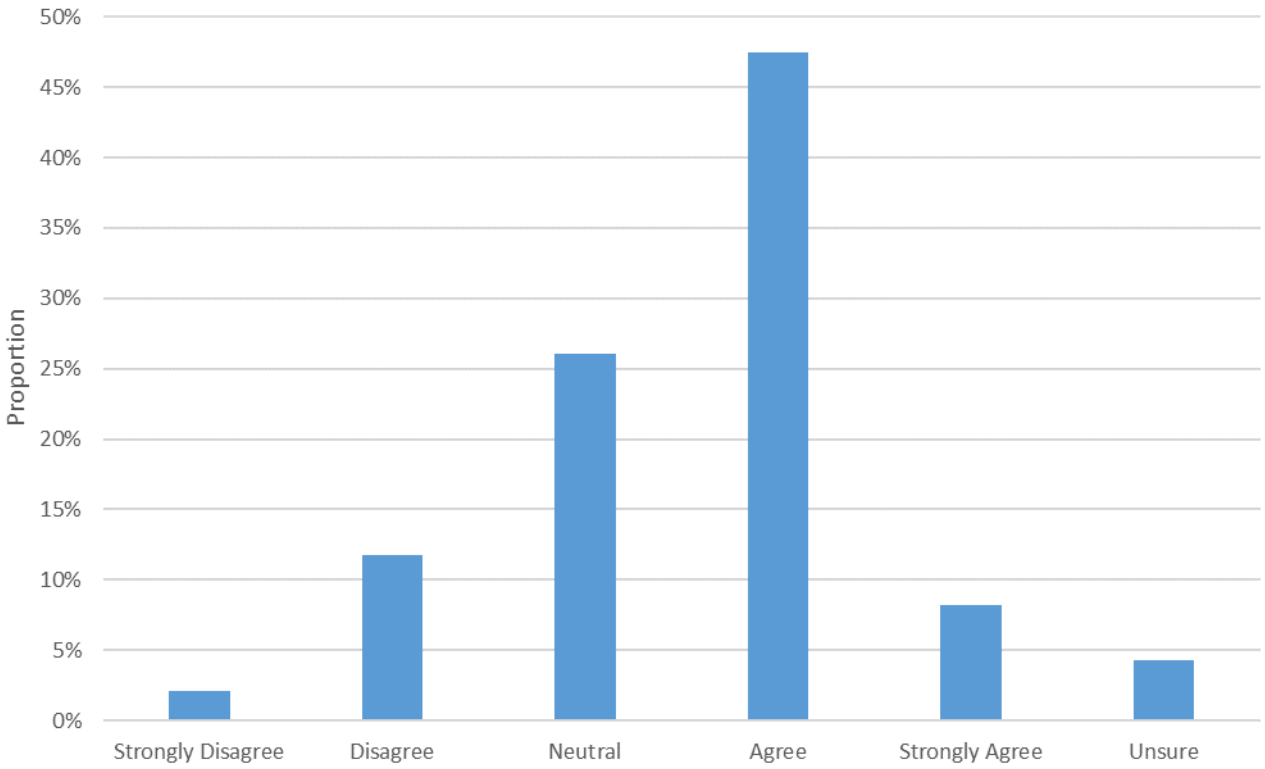
Q22 provided peer reviewers with the opportunity to provide comments on why they were unable to undertake consistent assessment of applications or suggestions for improvements to increase assessment consistency. Of the peer reviewer respondents, 18.9% (53 of 281) provided a comment in response to this question.

Of the peer reviewer respondents, 20.8% (11 of 53) commented on the impact type selection by applicants. These comments focused on all or most applicants picking the same impact types, applicants picking either one or two impact types, most applicants picking knowledge impact and a lack of applicants describing either economic or social impact.

Of the peer reviewer respondents, 26.4% (14 of 53) commented on aspects related to scoring of assessments, commonly noting the following:

- The difficulty of comparing applications, including comparing between basic researchers and clinicians 9.4% (5 of 53)
- The difficulty of assessing when either multiple impacts are selected or comparisons when applicants select different numbers of impacts 9.4% (5 of 53)
- The difficulty of assessing knowledge impact or impacts outside the area of peer reviewer expertise, 3.8% (2 of 53).

Figure 25. **Peer reviewer responses to Q21: "I was able to deliver consistent assessment of applications across the 4 impact types:"**



Of the peer reviewer respondents, 18.9% (10 of 53) suggested that the way that the impact types were addressed by applicants reflected confusion by the applicants. Peer reviewers felt that applicants had difficulty distinguishing impact types, confused impact types or picked the wrong impact type.

An additional 18.9% (10 of 53) of peer reviewer respondents reflected on perceived biases within the system. Comments were made on the perceived devaluation of knowledge impact and disadvantages to basic science (6 of 10). Other comments were made on the impact types being weighted towards certain types of research activity and that some impact areas were more easily quantifiable.

Respondents provided suggestions for improvements on the structure of the section (8 of 53) or guidance for applicants and peer reviewers (3 of 53). These included:

- The removal of impact type selection (3 of 53)
- Increased space for applicant responses (2 of 53)
- More guidance needed for applicants as applicant confusion makes assessment difficult (2 of 53)
- Better examples needed of impact types.

Findings

While leadership level continues to exert an effect upon impact scoring even after implementation of the revised collection of R2O information and SofE guidance, the scoring differences among the

leadership levels have been reduced. Moreover, because the EL1, EL2 and L grants are funded from separate budgets, these impact scoring differences do not ultimately influence funded rates at the different levels, even though all three levels are assessed together.

The number of impact types selected by applicants is somewhat predictable from their leadership level – with more experienced applicants claiming a greater number of impacts. That said, the fact that L3 applicants are less likely to claim knowledge impacts when they are more likely to have produced them also indicates that impact type is an unreliable source of information about the research. Some peer reviewers also stated this in their comments.

Peer reviewers both claim that they were able to consistently assess applications across the impact types and appear to have scored different impact types very similarly.

Insufficient evidence exists to understand whether scoring differences appear between applications claiming smaller and larger numbers of impacts.

Summary of findings

The evaluation questions and findings are shown below:

A1 – How easy was it to provide the impact text?

Applicants found it difficult to provide the impact text and both applicants and peer reviewers thought that there was substantial overlap in the text provided in the three impact sub-sections. Applicants found the examples of evidence to support their impact statements provided by NHMRC helpful but, thought that they could be improved.

STATUS – Review required

A2 – Are the impact types useful?

Most applicants consider that the four impact types, knowledge, health, social and economic, allow them to report all the impacts that they would like to, however some peer reviewers raised concerns that applicants had selected incorrect impact types.

Data analysis confirms that, across all applications, the impact types selected do relate to the impact text being provided, however applicants report knowledge impacts even if they have not selected this impact type.

STATUS – Review required

A3 – What types of impacts are being reported?

Applicants more often report on knowledge impacts than they do on benefits experienced by stakeholders beyond the research sector.

STATUS – Review required

A4 – Is the impact text duplicating the publications text?

The impact and publications texts provided by applicants overlap substantially.

STATUS – Review required

B1 – How easy was it to assess the impact text?

Peer reviewers expressed mixed views about how easy the impact text was to assess but were more united in finding the Category Descriptors to be unhelpful and offered a variety of suggestions for their improvement.

STATUS – Review required

B2 – Do applicant characteristics influence scoring of impact?

Applicant characteristics do not appear to be associated with differences in scoring.

STATUS – Functioning adequately

B3 – What factors affect impact scoring?

Since the introduction of the collection of more structured R2O information and the revised SofE, Leadership level has become less predictive of the impact score. There is no apparent relationship between impact type and impact score.

STATUS – Functioning adequately

Discussion

The evaluation described above was a process evaluation, ultimately intended to determine whether implementation of the RITRA framework has led to consideration of past research impact in peer review of researcher track record.

While noting the short time frames since RITRA was first implemented and that applicants and peer reviewers are still learning how best to engage with the RITRA framework, the evaluation has identified several issues arising from the framework implementation that need to be addressed. These include:

- duplication of information provided by applicants in the three impact subsections and between the impact and publications texts.
- difficultly experienced by applicants, and especially those early in their career, in providing retrospective impacts
- possible confusion about the nature of research impact
- difficulty experienced by peer reviewers in assessing the information provided in the applications
- insufficient guidance provided by the category descriptors.

Based on these findings, the RITRA framework could be improved to ensure it is achieving its short-term outcomes and overall objectives.

Recommendations

In order to address the issues identified by this evaluation, some revisions could be made to the RITRA framework, as follows:

- To reduce duplication of content in the application, the three components of the Research Impact section could be combined into one or two components
- To provide more space for applicants to explain their research impact, all evidence for impact – in the form of URLs or document citations – could be included within a separate free-text field of the application form
- To align with most applicants describing the generation of knowledge that may lead to impact, applicants could instead outline their pathway(s) to impact and engagement with research end users. NHMRC could develop advice about pathways to impact and describe various markers for each impact pathway type
- To help peer reviewers assess the applicant's impact pathway, the Category Descriptors could be revised – ideally assisted by cognitive interviews – to ensure that applicants, peer reviewers and NHMRC all understand them in the same way.

Revision of the RITRA framework should be guided by an expert working group.

Appendices

Appendix A – RITRA framework logic model

Objectives

- Include consideration of past research impact in peer review of track record
- Encourage consideration of future impact in the planning and conduct of NHMRC-funded research
- Increase the translation of NHMRC-funded research

Inputs

- Investigator and Synergy Grant schemes
- Grant Opportunity
- Guide to Applicants
- Peer Review Guidelines

Activities

Component

- Development of a track record assessment framework that includes publications, impact and leadership
- Development of guidance for applicants on how to prepare a grant application including description of past research impacts
- Development of guidance for peer reviewers on how to assess descriptions of past research impacts

Participation

- ONHMRC staff
- Track Record Working Group
- NHMRC Principal Committees

Outcomes

Short term (Evaluation phase 1)

- Information about past research impacts including in grant applications
- Past research impacts assessed by peer reviewers

Medium term (Evaluation phase 2)

- Increased consideration of future impact by researchers when planning and conducting research
- Increased support by Administering Institutions for researchers to contribute to future impacts

Long term

- Increased translation of NHMRC-funded research.

Assumptions

Applicants are able to provide impact information. Peer reviewers are able to assess this information

External factors

Broader context, including domestic and international government policies, programs, initiatives and funding schemes that encourage or discourage a focus on impact, or that have an impact on NHMRC and/or the research sector relevant to impact.

Appendix B – Examples of evidence

Impact type
Knowledge impact Description of research impact New knowledge demonstrating the benefits emerging from adoption, adaption or use of new knowledge to inform further research, and/or understanding of what is effective. Examples of evidence (not exhaustive) • recognition of research publications (for example, citation metrics, particularly field weighted) • data sharing • contribution to registries or biobanks • prizes and conference presentations • uptake of research tools and techniques • evidence of uptake of the research by other disciplines.
Health impact Description of research impact Improvements in health through new therapeutics, diagnostics, disease prevention or changes in behaviour; or improvements in disease prevention, diagnosis and treatment, management of health problems, health policy, health systems, and quality of life. Examples of evidence (not exhaustive) • policy or program adopted • a clinical guideline adopted • international or national practice standards adopted • improved service effectiveness • Phase I, Phase II and Phase III clinical trials underway or completed • improved productivity due to research innovations (for example, reduced illness, injury) • Quality-Adjusted Life Years, Disability-Adjusted Life Years, Potential Years of Life Lost, Patient Reported Outcome Measures and other relevant indicators • relative stay index for multi-day stay patients, hospital standardised mortality ratio, cost per weighted separation and total case weighted separation reports (including community and government).
Social impact Description of research impact Improvements in the health of society, including the well-being of the end user and the community. This may include improved ability to access health care services; to participate socially (including empowerment and participation in decision making) and to quantify improvements in the health of society. Examples of evidence (not exhaustive) • uptake or demonstrated use of evidence by decision makers/policy makers • qualitative measures demonstrating changes in behaviours, attitudes, improved social equity, inclusion or cohesion • improved environmental determinants of health • improved social determinants of health • changes to health risk factors.

Appendix C – Charts of some RITRA dataset variables

The following charts describe some of the variables included in the RITRA dataset.

Figure 1. Number of applications in each Application year, 2019 to 2022

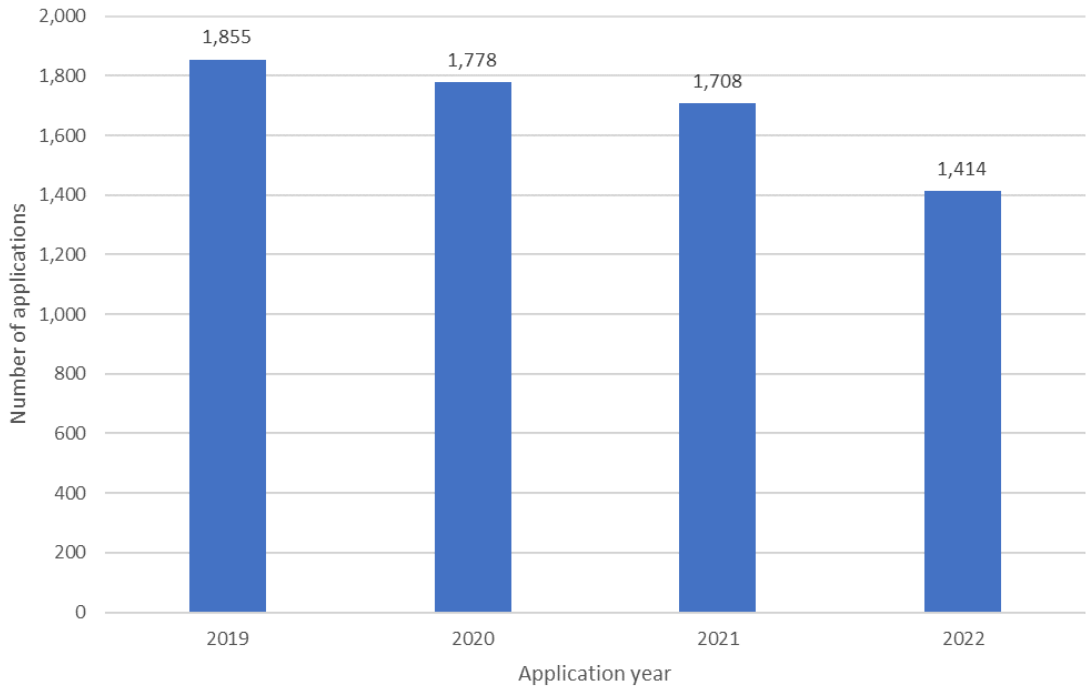


Table 1. Number of applications in each Application year, 2019 to 2022

Year	Count
2019	1,855
2020	1,778
2021	1,708
2022	1,414

Figure 2. Proportion of 2019-2022 applications in each Broad Research Area

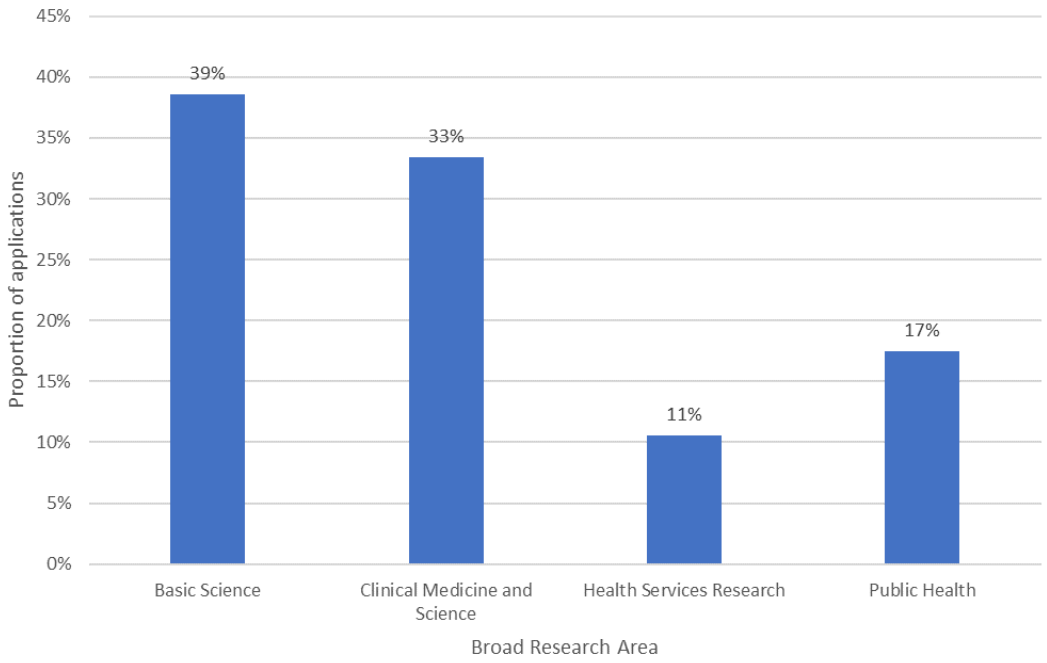


Table 2. Proportion of 2019-2022 applications in each Broad Research Area

Broad Research Area	Proportion
Basic Science	39%
Clinical Medicine and Science	33%
Health Services Research	11%
Public Health	17%

Figure 3. Proportion of applications indicating Indigenous research

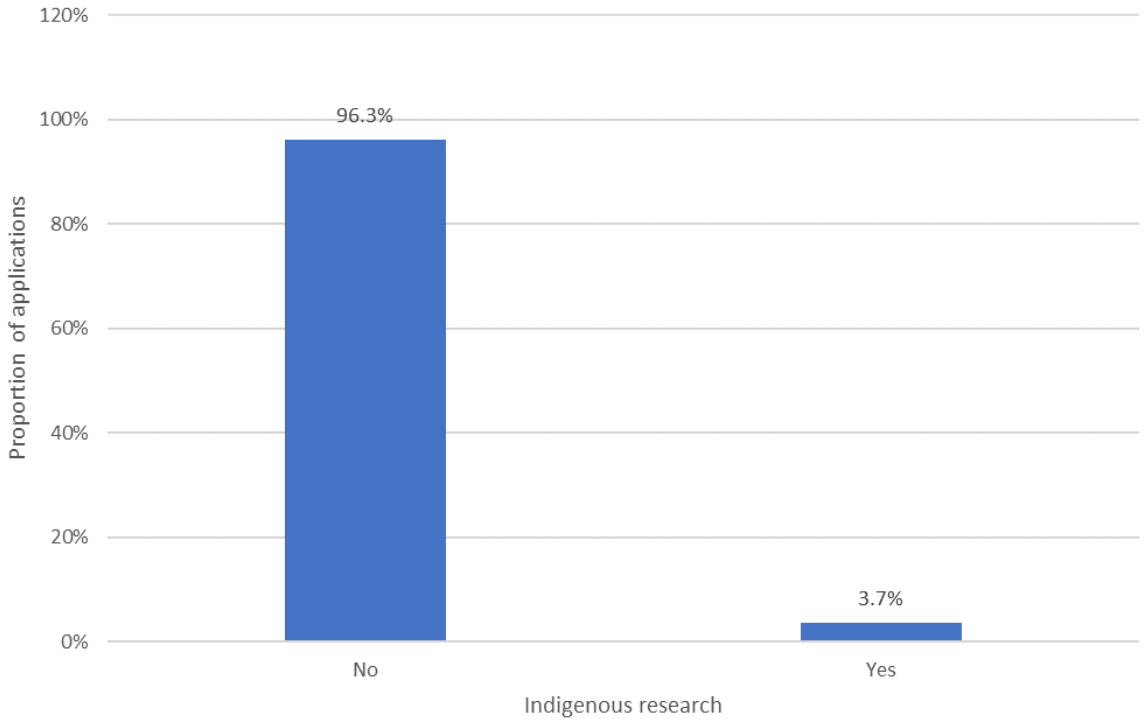


Table 3. Proportion of applications indicating Indigenous research

Indigenous research?	Proportion
No	96.3%
Yes	3.7%

Figure 4. Number of applications by applicant gender

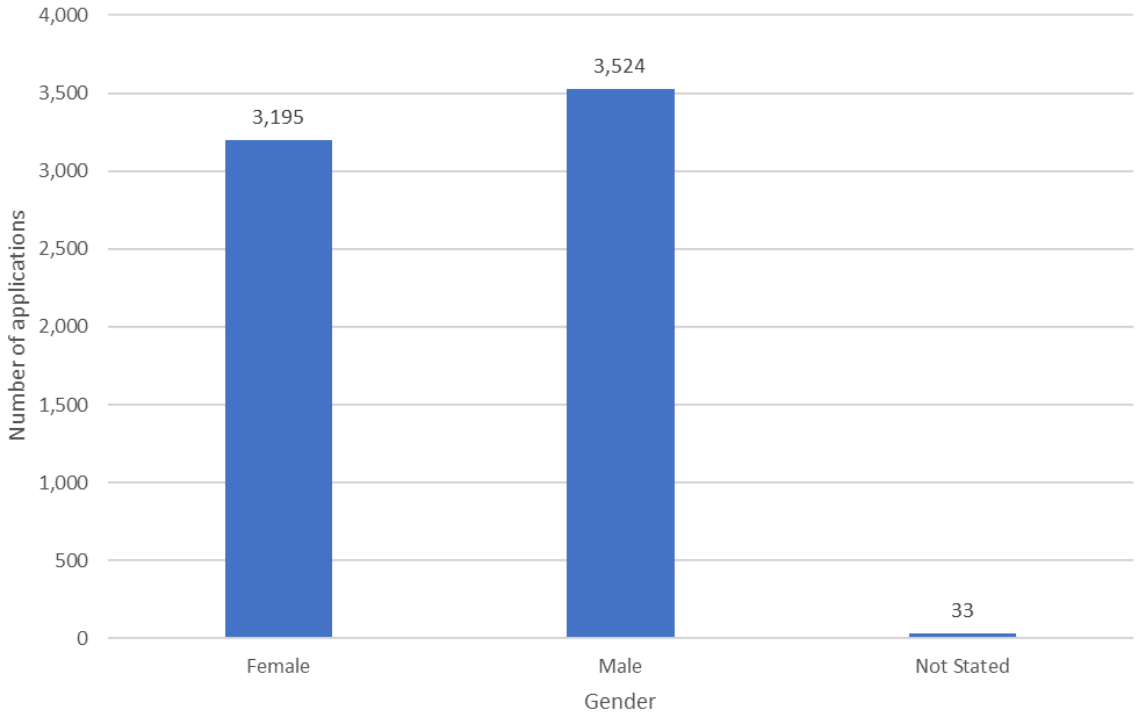


Table 4. Number of applications by applicant gender

Gender	Number
Female	3,195
Male	3,524
Not Stated	33

Figure 5. Number of applications by Application year (2019-2022) and Leadership level

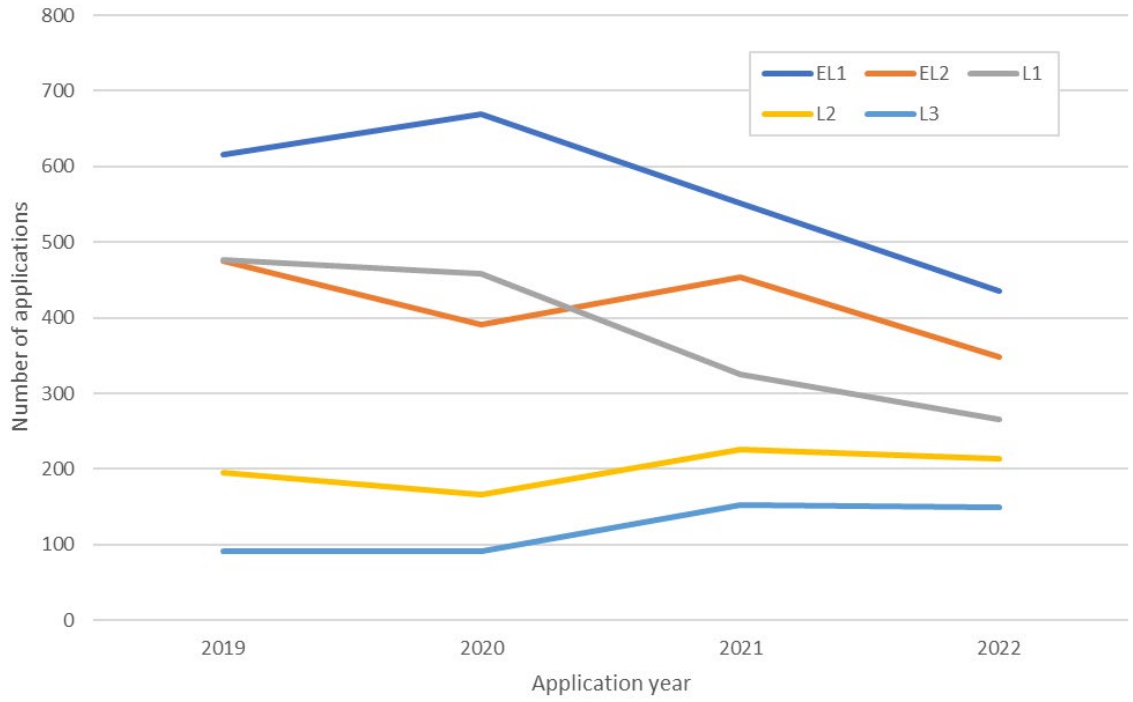


Table 5. Number of applications by Application year (2019-2022) and Leadership level

Year	EL1	EL2	L1	L2	L3
2019	616	475	477	195	92
2020	669	391	459	167	92
2021	551	454	325	226	152
2022	436	349	265	214	150

Figure 6. Proportion of applications claiming each Impact type by Application year (2019-2022)

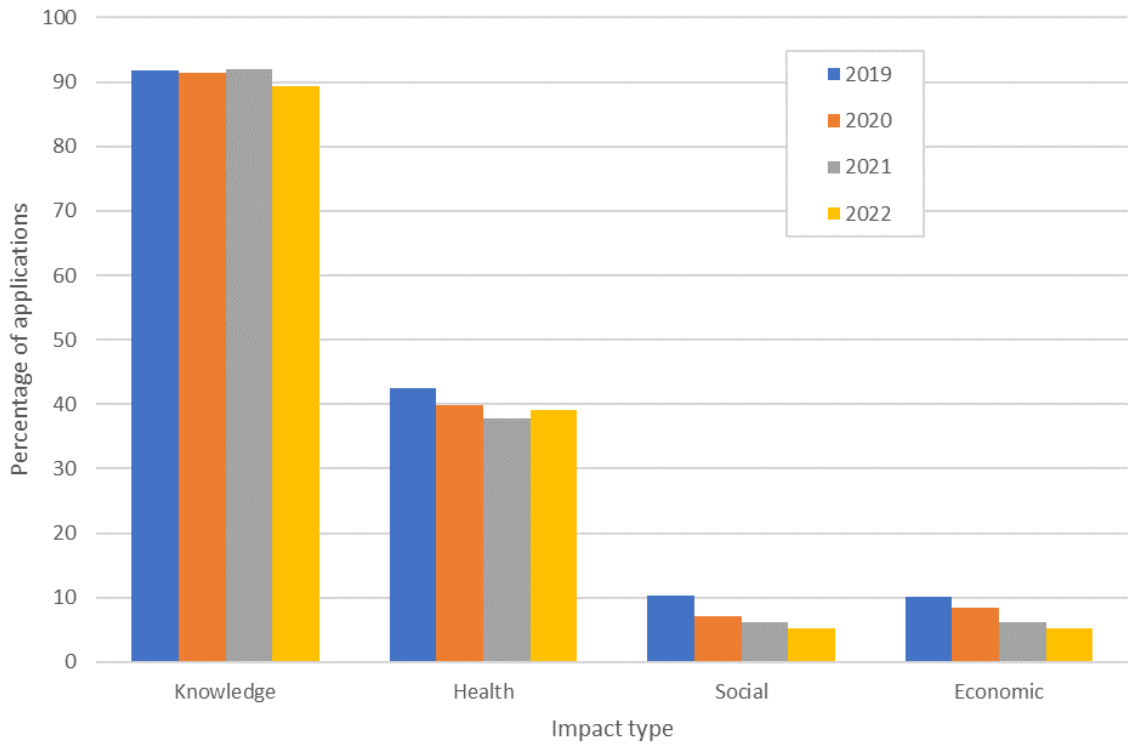


Table 6. Proportion of applications claiming each Impact type by Application year (2019-2022)

Year	Knowledge	Health	Social	Economic
2019	92	42	10	10
2020	91	40	7	8
2021	92	38	6	6
2022	89	39	5	5

Figure 7. Number of applications by number of Impact types and Application year (2019-2022)

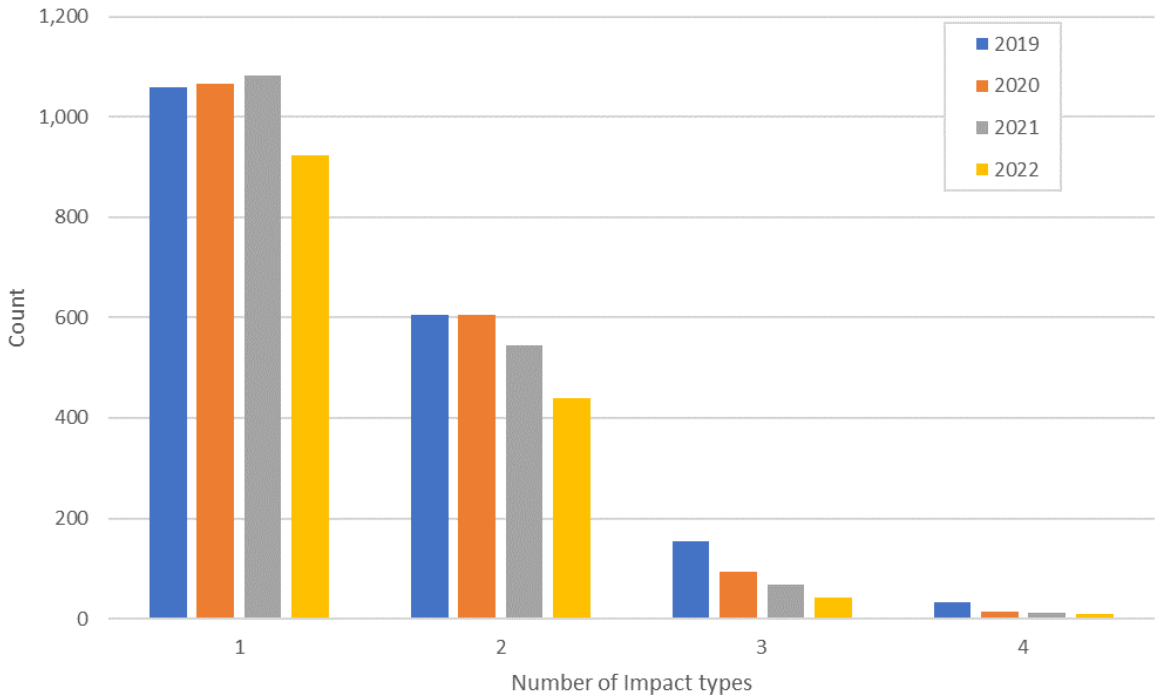


Table 7. Number of applications by number of Impact types and Application year (2019-2022)

Year	1	2	3	4
2019	1,060	605	154	34
2020	1,066	605	93	14
2021	1,082	544	69	13
2022	923	440	42	9

Figure 8. Mean peer reviewer score by assessment component and Application year (2019-2022)

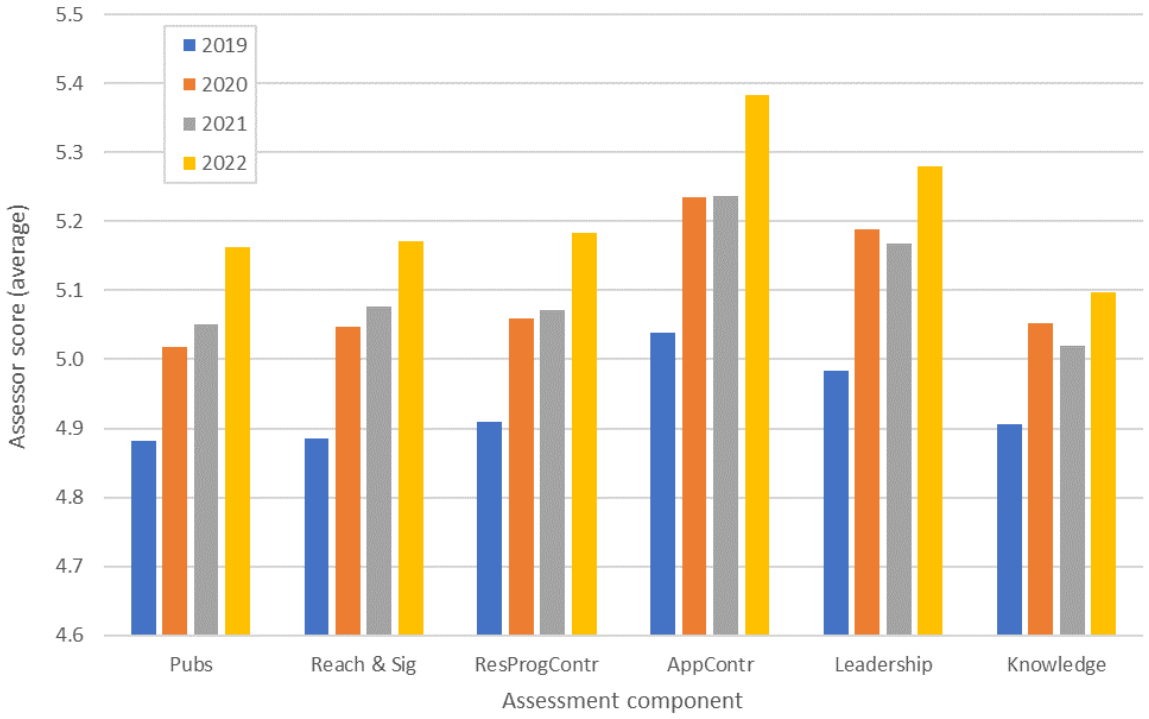


Table 8. Mean peer reviewer score by assessment component and Application year (2019-2022)

Year	Pubs	Reach & Sig	ResProgContr	AppContr	Leadership	Knowledge
2019	4.9	4.9	4.9	5.0	5.0	4.9
2020	5.0	5.0	5.1	5.2	5.2	5.1
2021	5.1	5.1	5.1	5.2	5.2	5.0
2022	5.2	5.2	5.2	5.4	5.3	5.1

Appendix D – Applicant survey

Q1. The four impact types allow me to report all the impacts I would like to.

Strongly agree – Agree – Neutral – Disagree – Strongly disagree – Unsure

---

Q2. Provide any comments on the definitions of the four impact types and/or any other impact types you would like included.

Q3. The examples of evidence were helpful in supporting the research impact components within my application.

Strongly agree – Agree – Neutral – Disagree – Strongly disagree – Unsure

---

Q4. Provide any further comments about the examples of evidence.

Q5. Filling in the following components is

Very easy – Easy - Neither difficult nor easy – Difficult - Very difficult – Unsure

- Reach and significance of the research impact
- Research program's contribution to the research impact
- Applicant's contribution to the research program

Q6. The information I provided in these 3 components overlapped:

Not at all – A little – Moderately – A lot – Unsure

Q7. Provide any additional comment on your thoughts on the 3 free-text field prompts that make up the research impact section.

Q8. The information I provided in the publications section overlapped with the information I provided within the research impact section:

Not at all – A little – Moderately – A lot – Unsure

Q9. How many of the publications listed in your top 10 publications were used as evidence in your research impact section?

0 – 1-2 – 3-4 – 5-6 – 7-8 – 9-10

Q10. Provide additional comment about your selection to the questions above.

Appendix E – Peer reviewer survey

Q15. How helpful were the category descriptors for each component, when assessing the applicants' research impact section?

Very helpful – Helpful – Neither – Unhelpful – Very unhelpful – Unsure

---

Q16. Provide any additional comment on the category descriptors for the 3 components of the research impact section, including any changes you would like to see to assist in scoring applications. (Word limit 5000)

Q19. Overall, how much did the information provided in the publications section overlap with the information provided in the research impact components?

Not at all – A little – Moderately – A lot – Unsure

Q20. Provide additional comment about your selection to the question above. (Word limit 2000)

Q21. To what extent do you agree with the following statements about the 4 types of impact?

- The 4 impact types cover the impacts described by applicants
- I was able to deliver consistent assessment of applications across the 4 impact types

Strongly agree – Agree – Neutral – Disagree – Strongly disagree – Unsure

---

Q22. If you selected Disagree or Strongly Disagree for the ability to undertake consistent assessment of application across assessments could you describe why? Do you have any suggested improvements or other comments? (Word limit 2000)

Q23. Provide any comments on the definitions of the 4 impact types and/or any other impact types you would like included. (Word limit 2000)

Q24. How easy was it to assess the following components?

- Reach and significance of the research impact
- Research program's contribution to the research impact
- Applicant's contribution to the research program

Very easy – Easy - Neither difficult nor easy – Difficult - Very difficult – Unsure

---

Q25. Overall, how much did the information provided by applicants within the 3 research impact components overlap?

Not at all – A little – Moderately – A lot – Unsure

Q26. When I am assessing the research impact section:

- I read all the text first and then score the components.
- I read the text of each component and score it before moving to the next.

Q27. Provide any additional comment on your thoughts on the research impact section. (Word limit 2000)

Appendix F – Topics in total impact text

#	Topic cluster	Topic keywords	Eigen value
1	Publications	AUTHOR; SENIOR; ST; PAPERS	4.86
2	Medical topic	PROTEIN; STRUCTURE; PROTEINS; FUNCTION; SIGNALLING; RECEPTOR; STRUCTURAL	4.22
3	Grants/ Researchers	NHMRC; CIA; GRANT; PROJECT; GRANTS; CIB; APP; GNT; FUNDED; CI; CRE; MRFF; IDEAS; ARC	3.76
4	Recognition	INVITED; KEYNOTE; SPEAKER; PLENARY; PRESENTATIONS; TALKS; CONFERENCE; CONFERENCES; INTERNATIONAL; SYMPOSIUM	3.5
5	Stakeholder	GOVERNMENT; NSW; STATE; HEALTH; SERVICE; SERVICES; VICTORIAN; QLD; VICTORIA; DEPARTMENT; WA; POLICY	3.21
6	Outcomes and impacts	COLLECTION; DATA; DESIGN; ETHICS; ANALYSIS; WRITING; RECRUITMENT; MANUSCRIPT; INTERPRETATION; RESPONSIBLE; ASPECTS; PROTOCOL	2.8
7	Publications	HTTPS; WWW; AU; ORG; GOV; PDF	2.67
8	Grants/ Researchers	CAPACITY; BUILDING	2.52
9	Recognition	AWARD; AWARDS; PRIZE; EXCELLENCE; AWARDED; TRAVEL; RECEIVED; WON; PRESTIGIOUS; YOUNG; RECOGNITION; OUTSTANDING; SOCIETY; RECOGNISED; PRESENTATION	2.42
10	Recognition	RADIO; ABC; MEDIA; INTERVIEWS; NEWS; COVERAGE; FEATURED; ONLINE	2.36
11	Publications	WEIGHTED; CITATION; AVERAGE; FIELD; INDEX; SCIVAL	2.28
12	Publications	NAT; REV; IMMUNOL; BIOL; MOL; GENET; NATURE; PNAS; CELL; IMMUNITY	2.23
13	Research methodology	II; III; PHASE; TRIALS; NCT; TRIAL	2.16
14	Medical topic	ANXIETY; DEPRESSION; DISORDERS; DISORDER; SYMPTOMS; PSYCHIATRIC; SCHIZOPHRENIA; COGNITIVE	2.08
15	Grants/ Researchers	STUDENTS; PHD; HONOURS; SUPERVISED; STUDENT; SUPERVISOR; SUPERVISION; STAFF; PRIMARY	2.03
16	Outcomes and impacts	INTELLECTUAL; INPUT; TECHNICAL; LEADERSHIP; PROVIDED	1.99
17	Outcomes and impacts	REDUCED; RATES; HIGHER; COMPARED; INCREASED; RATE; SIGNIFICANTLY; MORTALITY; SHOWED; IMPROVED	1.96
18	Recognition	MEETING; ANNUAL; SOCIETY; AMERICAN; CONGRESS; ASSOCIATION; SCIENTIFIC; EUROPEAN; SYMPOSIUM; PRESENTATION; CONFERENCE; ZEALAND	1.91
19	Outcomes and impacts	ACROSS; COUNTRIES; DISCIPLINES; AREAS; FIELDS; SUBJECT; INSTITUTIONS; DIVERSE; MULTIPLE	1.89
20	Medical topic	GENETIC; GENE; VARIANTS; GENES; EXPRESSION; MUTATIONS	1.87
21	Publications	BOOK; CHAPTERS; CHAPTER; REVIEWS; WRITE	1.85

22	Publications	SCHOLAR; GOOGLE; CITATIONS; INDEX	1.83
23	Outcomes and impacts	DEVELOPED; TOOLS; METHODS; TOOL; ASSESSMENT; SOFTWARE; VALIDATED	1.8
24	Medical topic	VACCINE; VACCINES; INFLUENZA; VACCINATION; MALARIA	1.76
25	Stakeholder	USA; UK; CANADA; GERMANY; CHINA; AUSTRALIA; EUROPE; NZ	1.75
27	Other	TERM; LONG	1.72
26	Outcomes and impacts	THERAPEUTIC; DRUG; TARGETS; DRUGS; TARGET; ANTI; POTENTIAL; TARGETING; TREATMENT	1.72
28	Outcomes and impacts	GUIDELINES; PRACTICE; CLINICAL; MANAGEMENT; RECOMMENDATIONS; POLICY; GUIDELINE; INFORMED; CHANGED; NATIONAL; CONSENSUS	1.7
29	Outcomes and impacts	OUTCOMES; PATIENT; IMPROVE; OUTCOME; CARE; IMPROVED; IMPROVING	1.69
30	Outcomes and impacts	KNOWLEDGE; RESEARCH; PROGRAM; IMPACT; SIGNIFICANT; IMPACTS; CONTRIBUTED; CONTRIBUTION; EXCEPTIONAL	1.67
31	Research methodology	SYSTEMATIC; REVIEW; REVIEWS; COCHRANE; META	1.66
32	Outcomes and impacts	MODELS; ANIMAL; MOUSE; VIVO; VITRO; MODEL; HUMAN; PRECLINICAL; MICE	1.64
33	Publications	PEER; REVIEWED	1.64
34	Medical topic	CANCER; BREAST; PROSTATE; CANCERS; TUMOUR	1.63
35	Medical topic	KIDNEY; CHRONIC; DISEASE; CKD; RENAL	1.61
36	Medical topic	OLDER; ADULTS; FALLS; PEOPLE; PREVENTION	1.6
38	Medical topic	INFECT; DIS; CLIN	1.59
37	Recognition	RECORD; TRACK; STRONG	1.59
39	Recognition	COMMITTEE; ADVISORY; MEMBER; GROUP; CHAIR; WORKING; EXPERT	1.58
40	Recognition	ALTMETRIC; SCORE; ATTENTION; OUTPUTS; TOP; MEDIA; ATTRACTED	1.57
41	Stakeholder	UNIVERSITY; MELBOURNE; MONASH; INSTITUTE; PROF; SCHOOL; CENTRE; SYDNEY; DR; MEDICAL; HOSPITAL; COLLABORATION	1.56
42	Outcomes and impacts	PARADIGM; SHIFT; CHANGING	1.55
43	Medical topic	RESISTANCE; ANTIMICROBIAL; ANTIBIOTIC; AMR; BACTERIAL; INFECTIONS	1.54
44	Stakeholder	ASIA; PACIFIC; EUROPE	1.54
45	Research methodology	COHORT; LARGE; STUDY; LONGITUDINAL; STUDIES; LARGEST; SCALE	1.53
46	Stakeholder	ABORIGINAL; INDIGENOUS; COMMUNITIES; AUSTRALIANS; COMMUNITY; FAMILIES; CHILDREN	1.52
48	Research methodology	RANDOMISED; CONTROLLED; TRIAL; TRIALS; RCT	1.51
47	Other	DECISION; MAKING	1.51

49	Grants/ Researchers	FUNDING; SECURED; COMPETITIVE; MILLION; GRANTS; SUPPORT; OBTAINED; CIA	1.5
50	Outcomes and impacts	COMMUNITY; TRAINING; PROFESSIONALS; WORKSHOPS; EDUCATION; PROFESSIONAL; CLINICIANS; PUBLIC; ENGAGEMENT	1.5
51	Medical topic	PHYSICAL; ACTIVITY; BEHAVIOUR	1.49
52	Publications	OPEN; BMJ	1.48
54	Outcomes and impacts	PATENT; WO; PATENTS; APPLICATIONS; APPLICATION	1.47
53	Publications	JOURNAL; JOURNALS; LANCET; PUBLISHED; PSYCHIATRY; JAMA; NEUROLOGY; HIGH	1.47
57	Grants/ Researchers	EARLY; CAREER; RESEARCHER; STAGE; FELLOWSHIP	1.46
55	Outcomes and impacts	PLAYED; ROLE; INSTRUMENTAL; KEY; ESTABLISHING; CENTRAL	1.46
56	Recognition	INVITATIONS; SPEAK; PRESENT; CONFERENCES; INTERNATIONAL; WRITE; MEETINGS; NATIONAL; INVITATION	1.46
58	Medical topic	HEART; FAILURE; CARDIAC; FOUNDATION; AF; AMERICAN	1.44
59	Outcomes and impacts	REACH; SIGNIFICANCE; IMPACT; EVIDENCED	1.44
60	Outcomes and impacts	COST; EFFECTIVENESS; EFFECTIVE; ECONOMIC; COSTS	1.44
61	Outcomes and impacts	EVIDENCE; CORROBORATING; BASE; PROVIDED; BASED	1.42
62	Medical topic	MECHANISMS; MOLECULAR; UNDERLYING; UNDERSTANDING; BIOLOGY; CELLULAR	1.41
63	Medical topic	STEM; CELL; CELLS	1.41
64	Medical topic	BRAIN; STIMULATION; INJURY; NEURAL	1.4
65	Medical topic	FOOD; DIETARY; DIET; NUTRITION; HEALTHY; OBESITY; MATERNAL; CHILD	1.4
67	Medical topic	RISK; FACTORS; FACTOR; CVD; PREVENTION	1.39
66	Research methodology	EXPERIMENTS; DESIGNED; WROTE; PERFORMED; CONCEIVED; MANUSCRIPTS	1.39
68	Research methodology	DIAGNOSTIC; TESTS; TESTING; DIAGNOSIS; TEST; CLINICAL	1.38
69	Stakeholder	COLLABORATIONS; INTERNATIONAL; MULTIDISCIPLINARY; COLLABORATIVE; TEAM; RESEARCHERS; ESTABLISHED; NETWORK; NATIONAL; INDUSTRY; CLINICIANS; PARTNERS; RESEARCH; SCIENTISTS	1.38
70	Stakeholder	INTERNATIONALLY; NATIONALLY; RECOGNISED	1.38
71	Medical topic	ASTHMA; SEVERE; RESPIRATORY	1.37
72	Publications	CITED; TIMES; HIGHLY; INFLUENTIAL; PAPER	1.37
73	Medical topic	PRESSURE; BLOOD; HYPERTENSION	1.36

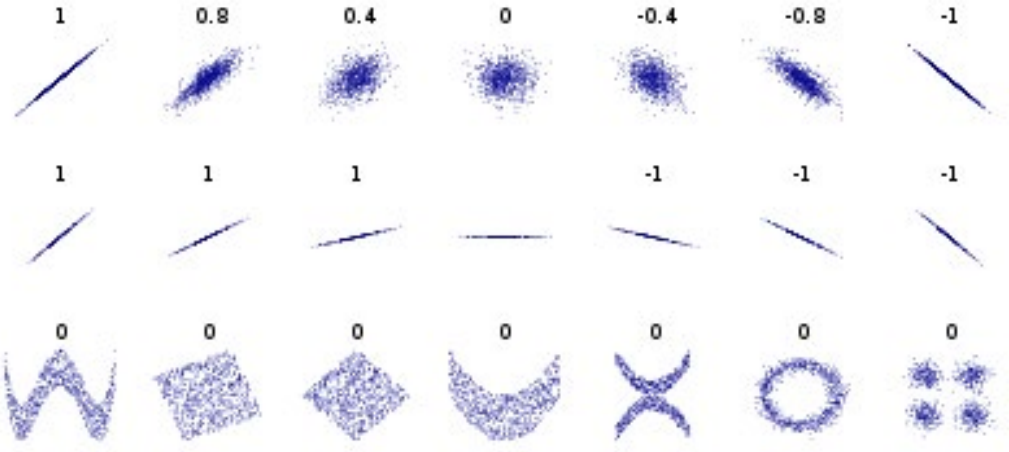
74	Outcomes and impacts	EXPERTISE; SKILLS	1.36
77	Grants/ Researchers	POST; DOCTORAL; YEARS; PHD; PAST	1.35
78	Grants/ Researchers	INVESTIGATOR; CHIEF; YOUNG; LEAD	1.35
76	Other	RELEVANT; CLINICALLY; QUESTIONS	1.35
75	Stakeholder	ROYAL; COLLEGE; AUSTRALASIAN; SOCIETY; HOSPITAL; AMERICAN; ZEALAND; AUSTRALIAN	1.35
79	Outcomes and impacts	ENGINEERING; BIOMEDICAL; MATERIALS; TISSUE; SCIENCE; CHEMISTRY; APPLICATIONS; SCIENCES; MEDICINE	1.34
80	Outcomes and impacts	PHARMACEUTICAL; COMPANIES; INDUSTRY; COMPANY; COMMERCIAL; PARTNERS	1.34
82	Medical topic	DIABETES; TYPE; INSULIN; METABOLIC; METABOLISM; COMPLICATIONS	1.33
83	Other	REAL; WORLD; TIME; LARGEST; CONGRESS	1.33
81	Outcomes and impacts	SAFETY; QUALITY; COMMISSION; AGED; CARE; MEDICATION; MEDICINES; ROYAL	1.33
86	Medical topic	PLAN; ACTION; STRATEGIC; MECHANISM; PREVENTION; SUICIDE	1.32
85	Other	SCI; REP; MED; INT; SOC	1.32
84	Publications	RANKED; TOP; WORLDWIDE; JOURNALS; SCIVAL; TOPIC; HIGHEST	1.32
87	Medical topic	IMMUNE; RESPONSES; INFECTION; VIRAL; IMMUNITY; HOST; ANTI; INFLAMMATORY; CELLS	1.31
88	Medical topic	ALZHEIMER; AD; DEMENTIA; DISEASE; AGEING; COGNITIVE	1.31
90	Medical topic	LOW; PAIN	1.31
89	Research methodology	META; ANALYSES; ANALYSIS; STATISTICAL; DATA	1.31
91	Medical topic	DISEASES; INFECTIOUS; DISEASE; INFLAMMATORY	1.3
94	Medical topic	WEIGHT; LOSS; GAIN; OBESITY; AF	1.3
93	Other	BASIS; FORMED	1.3
92	Publications	FWCI; CITES	1.3
95	Grants/ Researchers	ARC; FELLOWSHIP; LINKAGE; FELLOWSHIPS; FUTURE; POSTDOCTORAL	1.29
96	Other	SCALE; LARGE	1.29
97	Recognition	ORAL; CONFERENCES; PRESENTATIONS; PRESENTED; CONFERENCE; PRESENTATION; INTERNATIONAL; NATIONAL; FINDINGS; SELECTED; TRAVEL	1.29
98	Medical topic	CLINICAL; SCIENCE; BASIC; TRIALS; TRANSLATIONAL	1.28
99	Medical topic	SEQUENCING; GENOME; RNA; DNA; SINGLE; GENE; GENOMICS; BIOINFORMATICS; EXPRESSION	1.28
100	Medical topic	MENTAL; HEALTH; ILLNESS; YOUTH; PEOPLE	1.28

Appendix G – Correlations and statistical significance

In statistics, a 'correlation' refers to the degree to which a pair of variables are linearly related, meaning that constant changes to the value of one variable are associated with constant (positive or negative) changes to the value of the other. As shown in Figure 1, calculation of a correlation coefficient between two variables only provides insight to the extent that those variables have a linear relationship. If they have other types of relationships these may be apparent in a scatterplot but not in a correlation.

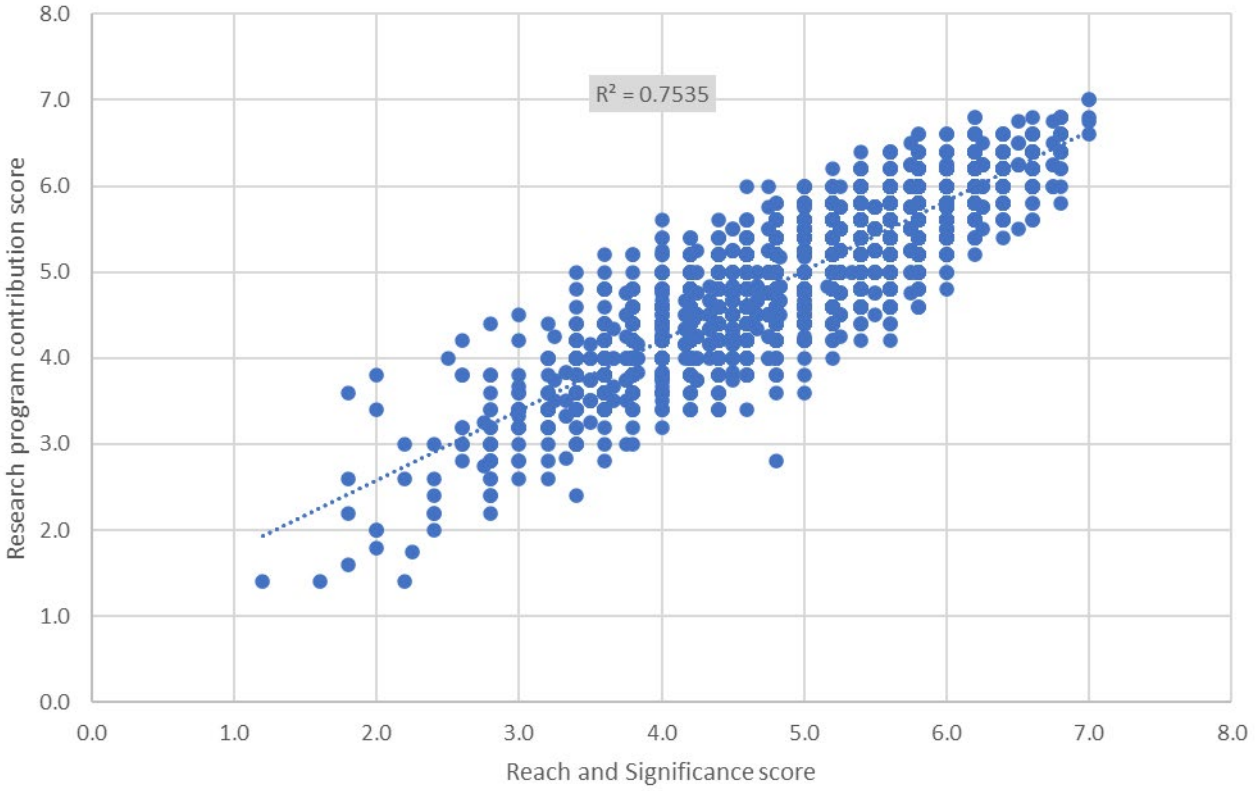
The example of a scatterplot provided in Figure 2 compares peer reviewer scores for two research impact components: reach and significance and research program contribution. Calculation of the correlation coefficient (Pearson's r) between these two variables produces a value of 0.868. As shown in Figure 1, the closer the r value is to 1 the closer the relationship between the variables is to being purely linear. Squaring the r value provides a new statistic (R^2) which is a measure of the proportion of the variability in one variable explained by values of the other variable. Where $R^2 = 0.75$, the values of one variable predict 75% of the variability in the values of the other variable.

Figure 1. **Types of relationships between two variables as revealed by scatterplots, and their associated correlation coefficients (r values)**



Credit: https://en.wikipedia.org/wiki/Correlation

Figure 2. **Scatterplot showing values of two research impact scores: Reach and Significance versus Research Program Contribution**



When r is calculated, the statistical significance (p value) of the r value is calculated at the same time. Throughout this evaluation, p <0.01 has been used as the standard p value. The choice of 0.01 was arbitrary. Had either of the two other customary p values (0.05, 0.001) been chosen the major outcomes of the analyses described below would have been fundamentally the same.

P values indicate the number of times that a sample collection of scores might be drawn from a larger population and possess, by chance alone, a certain magnitude of difference in mean value from that of the total population. p=<0.05 indicates 'less than five times per hundred'. p<0.01 indicates 'less than 1 time per hundred'. p<0.001 indicates 'less than one time per thousand'.

The lower the p value the more likely it is that the sample was drawn from some other population (i.e., that a genuine difference exists between the variables being compared). However, when the sample size is large (as it is in the case of the present evaluation, containing over 6,500 data points), even very small values of r might still be statistically significant. This is because large sample sizes amplify the detection of even very small differences.

Since a difference may statistically significant but still be too small to have any practical significance, within the present evaluation and following Cohen (1992)[22], values of Pearson's r:

- less than 0.1 are considered too low to be meaningful
- from 0.1 to 0.3 are considered small
- from 0.3 to 0.5 are consider medium

greater than 0.5 are considered large.

[22] Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155

Appendix H – Correlation matrix for profile variables

	1. AY	2.BS	3.CM	4.HS	5.PH	6.In	7.Gn	8.YP	9.CD	10.LL	11.Kn	12.He	13.So	14.Ec
1. AY	1													
2. BS		1												
3. CM			1											
4. HS				1										
5. PH					1									
6. In		-0.138			0.149	1								
7. Gn		-0.151		0.148	0.134		1							
8. YP							-0.162	1						
9. CD		-0.118					0.348		1					
10. LL							-0.197	0.862		1				
11. Kn		0.174		-0.109	-0.128						1			
12. He		-0.402	0.228	0.168						0.171	-0.288	1		
13. So		-0.194		0.124	0.259	0.122					-0.167	0.1	1	
14. Ec										0.105		0.119	0.107	1
15. IT		0.299	0.14	0.136	0.1					0.151				