

# Open Reforms Pilots: Spring 2013 Fellowship Competition Report

January 31, 2014

## Table of Contents

|   |    |
|---|----|
| Introduction .....  | 3  |
| Competition Overview .....  | 3  |
| Methods .....   | 4  |
| Summary of Results .....  | 4  |
| Conclusion and summary of changes as a results of the Pilot ..... | 5  |
| Appendix 1: Detailed Survey Results .....                         | 7  |
| Reviewing applications.....                                       | 11 |
| Peer Review Process & Technology .....                            | 12 |
| Appendix 2: Fellowship Reviewer Survey .....                      | 13 |

## Introduction

The success of the transition to the new Open Suite of Programs and peer review processes rests largely on our ability to efficiently and effectively pilot and test the functionality of peer review design elements in advance of their full implementation. As we complete each of the pilot studies, it is our intention to make findings available to contribute to the body of literature on peer review and program design.

Piloting peer review design elements will allow CIHR to adjust and refine processes and systems in order to best support applicants and reviewers. In order to identify areas for improvement, CIHR is collecting feedback from applicants, CIHR staff institutions, and peer reviewers through surveys.

It is important to note that when conducting pilots within existing competitions, CIHR will ensure that the quality of the peer review process is not compromised and that every application receives a fair, transparent and high-quality review.

Note: CIHR is also using data collected from this pilot to validate the optimal number of reviewers to assign to an application; and how often reviewers modify their reviews after viewing other reviewer's reviews/participating in the asynchronous online discussion. These analyses are being conducted separately and the results will be made available at a later date.

## Competition Overview

The Fellowship program was chosen for this pilot as it has been using a remote review process for many years as well as for the high volume of applications it receives. Therefore, the functionality implemented within this competition not only tested peer review design elements but also represented an enhancement to the existing peer review process. A detailed description of this process can be found [here](#). For this competition, 608 applications were reviewed by 209 reviewers

Within this first pilot, a number of program design elements remained unchanged:

1. Funds available
2. Evaluation criteria
3. Application-reviewer assignment process
4. Funding decision process
5. Rating scale
6. Application structure

The objectives of the pilot were to test proposed functionalities and processes related to remote review and asynchronous on-line discussion. The following design elements of the proposed Project Scheme were tested:

1. Increase the number of reviewers per application (from 2 to 3)
2. Remote review functionality (submitting preliminary and final reviews remotely)
3. Asynchronous online discussion

Note: The quality of individual reviews was not assessed during this pilot.

**Table 1: Design Elements from the Project Scheme tested in the Fellowship Phase 1 pilot**

| Design Element   | Objective  | Pilot Description  |
|--|--|--|
| Increased number of reviewers per application          | <ul style="list-style-type: none"> <li>• Improve reliability of results</li> <li>• Eliminate the current re-review process</li> <li>• Determine the optimal number of reviewers per application</li> </ul>   | <ul style="list-style-type: none"> <li>• Increase the number of reviewers per application from 2 to 3</li> </ul>   |
| Remote review - submitting preliminary & final reviews | <ul style="list-style-type: none"> <li>• Validate the proposed process</li> <li>• Improve consistency between reviewers</li> <li>• Monitor and collect data on all of the review components of the proposed design</li> <li>• Analyse trends if viewing the other reviewers reviews and how the discussion has influenced the reviewer's rating</li> </ul> | <ul style="list-style-type: none"> <li>• Reviewers submitted preliminary reviews and then had access to other reviewers' ratings and written reviews.</li> <li>• Reviewers were able to discuss applications/reviews, as appropriate, through an asynchronous online discussion tool.</li> </ul> |
| Asynchronous online discussion                         | <ul style="list-style-type: none"> <li>• Improve consistency between reviewers</li> <li>• Eliminate the current re-review process</li> <li>• Validate the proposed process</li> <li>• Monitor and collect data on all of the review components of the proposed design</li> </ul>   | <ul style="list-style-type: none"> <li>• The asynchronous online discussion allowed reviewers to consult and view each other's ratings and reviews and to discuss any areas of disagreement prior to submitting their final reviews.</li> </ul>  |

It is important to note the limitations to this pilot. The Fellowships program is a training awards program and therefore there are some notable program design differences when compared to a grant program. Notably, applications are assigned to one of five pools based on the candidate's background and their proposed research project. Reviewers appointed to each pool are then assigned randomly to applications within their respective pool. This is different than the more comprehensive expertise based assignment done in most CIHR grant programs.

### Methods

Peer review was conducted from April 2<sup>nd</sup>, 2013 to May 10<sup>th</sup>, 2013 by 209 reviewers. Reviewers had three weeks to participate in the asynchronous online discussion prior to submitting their final reviews. On May 15<sup>th</sup>, 2013, following the closure of the review process, an invitation to complete a survey was sent directly to the 209 reviewers using the Fluid Survey Invitation tool. It provided a URL specific to each participant, allowing them to stop, save and return to the survey at any time. CIHR monitored the response rates and two reminder emails were sent to ensure that the target response rate was achieved.

Applicants were not surveyed in this pilot as the enhancements only impacted reviewers. Applicants were surveyed for the second Fellowship pilot in fall 2013, as enhancements to the structured application were implemented.

### Summary of Results

The objective of Survey 1 (Appendix 2: Fellowships Reviewer Survey) was to assess the enhanced remote review process in ResearchNet by the reviewers with respect to:

- i. Their ResearchNet experience;
- ii. The usefulness of accessing other reviewer's reviews, and;
- iii. The usefulness of the asynchronous online discussion Tool.

The detailed survey results and summary of key findings and observations can be found in Appendix 1.

The survey response rate of 70% exceeded targets. However, the small overall number of reviewers will limit the statistical analysis that is possible. We will continue to build the dataset with subsequent pilots and surveys. Nevertheless, the overall results of this survey are positive within the context of this competition. A number of areas for improvements were identified by survey respondents and many of them have already been implemented within the subsequent pilots.

#### **i. ResearchNet Experience:**

The overwhelming majority (91%) of reviewers found the ResearchNet technology user friendly although recommendations were presented to enhance the system's interface. Overall, 84% of respondents agreed that the peer review process was efficient. There were a few technical problems and a number of suggestions for improvement were provided. Many of these improvements have been implemented within subsequent pilots (Fellowship Pilot Phase 2 & Knowledge Synthesis Pilot) as noted within the conclusion and summary of changes sections of this report.

#### **ii. Accessing other peer reviewers reviews:**

A majority of reviewers (89%) viewed the reviews submitted by other reviewers, and based on the comments provided to CIHR, they felt this functionality was very useful. On average, reviewers spent 2 hours or less reading the other reviews. However, there is currently no strong evidence to suggest that reading the other's reviews helped reviewers finalize their submitted score. Only 15% of the respondents said that the other reviews helped them finalize their submitted score often, very often or always. CIHR will explore the inconsistency between these two statements further through analysis of competition data, future pilots and will also refine the relevant questions within subsequent surveys.

#### **iii. Asynchronous online discussion:**

The asynchronous online discussion was used by the majority of reviewers (88%). Scoring discrepancy was the most common factor (87%) used to determine if an application required discussion. A minority of reviewers (29%) commented that direction from CIHR on whether or not to discuss an application would reduce their workload; however, 61% of reviewers felt that it should be at the reviewers' discretion whether or not to initiate a discussion. CIHR is currently exploring options to moderate the discussions to facilitate this process.

### **Conclusion and summary of changes as a results of the Pilot**

The enhancements to the remote review process within this pilot represented many positive changes to the review process and will serve as the foundation for subsequent pilots. In response to the data analysis, observation and reviewer feedback in this first pilot, enhancements were made to some of the processes and functionalities for subsequent pilots. To improve the effectiveness of the new adjudication process, several new features, listed below, were added to ResearchNet for the Knowledge Synthesis Grant and Fellowship Phase 2 pilots:

1. Deadline dates to complete the various task within the adjudication process are now displayed on ResearchNet.
2. The asynchronous online discussion tool and process will be improved by:
  - a. Allowing reviewers to access application and review information without closing the discussion tool;
    - i. A table will display the ratings from each reviewer at the top of the discussion screen;
  - b. Limiting the indentation of the text box to prevent the text from collapsing, regardless of the number of replies;
  - c. Opening the discussion period once all reviewers assigned to a particular application have submitted their preliminary reviews **or** by a pre-assigned date in the system, whichever comes earliest;
  - d. Shortening the timeframe for discussion;
  - e. Increasing accountability within the process by revealing the identities of all reviewers at the preliminary review due date, regardless if they have submitted or not;
    - i. Reviewers who have not submitted their review will see the identity of the other reviewers of the application but will not see their reviews;
3. The reviewer interface on ResearchNet will be improved by:
  - a. Displaying all of a reviewer's assigned applications on one page;
  - b. Providing reviewers with the functionality to expand the application details and view the other reviewers assessment;
  - c. Displaying ratings based on their actual value and not the calculated weighted value;
  - d. Creating a dynamic application rank order that will move applications up or down the rank list as ratings are entered for each adjudication criteria;
4. Written comments will be made mandatory as part of the structured review template.

## Appendix 1: Detailed Survey Results

### Response Rate

Survey was sent to 209 reviewers and 145 responses were received resulting in a response rate of 70%.

### Reviewer Characteristics

The reviewers were comprised of Mid-Career (45%) and Senior (42%) Scientists primarily from the Biomedical research domain (79%). Half of the respondents are Associate professors (50%) while 34% are Professors.

Note: the distribution (by pillar) of reviewers within this competition is representative of the distribution (by pillar) of the applications.

**Figure 1. Would you classify yourself as a:**

| Response   | Chart   | Percentage | Count      |
|--|---|------------|------------|
| Early career scientist (<5 years as an independent researcher) |    | 14%        | 20         |
| Mid-career scientist (5-10 years as an independent researcher) |   | 45%        | 65         |
| Senior scientist (>10 years as an independent researcher)      |  | 42%        | 60         |
| <b>Total Responses</b>   |   |            | <b>145</b> |

**Figure 2. Which research position(s) do you currently hold at this time?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Professor              |  | 34%        | 49         |
| Associate Professor    |  | 50%        | 72         |
| Assistant Professor    |  | 14%        | 20         |
| Researcher             |  | 10%        | 14         |
| Clinician              |  | 4%         | 6          |
| Other (specify):       |  | 2%         | 3          |
| <b>Total Responses</b> |   |            | <b>145</b> |

**Figure 3. Which of the following is your primary research domain:**

| Response  | Chart   | Percentage | Count      |
|---|---|------------|------------|
| Biomedical  |  | 79%        | 115        |
| Clinical  |  | 10%        | 14         |
| Health systems/services                               |  | 4%         | 6          |
| Social, cultural, environmental and population health |  | 7%         | 10         |
| <b>Total Responses</b>                                |   |            | <b>145</b> |

**Accessing the other reviewers reviews**

89% of the reviewers took the extra time to read the other reviewer’s reviews. Most (93%) spent a total of 2 hours or less reading the other reviews. However, there is no strong evidence to suggest that reading the other’s reviews helped finalize the reviewer’s submitted score. See figure 6.

**Figure 4. Did you look at the other reviews of the applications assigned to you?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Yes                    |  | 89%        | 129        |
| No                     |  | 11%        | 16         |
| <b>Total Responses</b> |   |            | <b>145</b> |

**Figure 5. How much additional time in total (all applications) did you spend reading reviews of other reviewers?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Under an hour          |  | 41%        | 53         |
| 1 to 2 hours           |  | 52%        | 67         |
| 3 hours or more        |  | 7%         | 9          |
| <b>Total Responses</b> |   |            | <b>129</b> |

**Figure 6. How often did the reading of other reviews help finalize your submitted score?**

| Response                   | Chart | Percentage | Count      |
|----------------------------|-------|------------|------------|
| Never (0%)                 |       | 8%         | 10         |
| Rarely (up to 25%)         |       | 29%        | 37         |
| Occasionally (>25% to 50%) |       | 48%        | 62         |
| Often (> 50% to 75%)       |       | 11%        | 14         |
| Very Often (>75%)          |       | 3%         | 4          |
| Always (100%)              |       | 2%         | 2          |
| <b>Total Responses</b>     |       |            | <b>129</b> |

### Online Discussion

The asynchronous online discussion was used by 88% of the reviewers. Scoring discrepancy was the common factor (87%) to determine if a discussion was needed. However, comments suggest that CIHR should provide more guidelines on when applications should be discussed. There was no clear indication if the discussion helped to modify their final scores. See figure 12.

### Representative Comments:

- Discuss application only when close to fundable.
- Some applications may not need discussion (scores are congruent amongst reviewers).
- CIHR should prompt/moderate discussions when necessary.
- Reviewers should NOT be able to submit final score without discussion.

**Figure 7. Did you initiate an online discussion thread?**

| Response               | Chart | Percentage | Count      |
|------------------------|-------|------------|------------|
| Yes                    |       | 72%        | 104        |
| No                     |       | 28%        | 40         |
| <b>Total Responses</b> |       |            | <b>144</b> |

**Figure 8. What factor(s) did you use to determine whether an online discussion was needed for an application?**

| Response                       | Chart   | Percentage | Count      |
|--------------------------------|---|------------|------------|
| Scoring discrepancy            |  | 87%        | 90         |
| Content clarification required |  | 2%         | 2          |
| Quality check                  |  | 3%         | 3          |
| Other:                         |  | 9%         | 9          |
| <b>Total Responses</b>         |   |            | <b>104</b> |

**Figure 9. Who should decide when an online discussion must take place?**

| Response  | Chart   | Percentage | Count      |
|---|---|------------|------------|
| CIHR (e.g., using a decision rule within ResearchNet) |    | 29%        | 42         |
| Reviewers (at their discretion)                       |    | 61%        | 87         |
| Don't know  |  | 10%        | 14         |
| <b>Total Responses</b>                                |   |            | <b>143</b> |

**Figure 10. Did you participate in an online discussion?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Yes                    |  | 88%        | 126        |
| No                     |  | 12%        | 17         |
| <b>Total Responses</b> |   |            | <b>143</b> |

**Figure 11. Did online discussion increase the amount of time required to complete the review of an application (on average) in this competition?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Under an hour          |  | 55%        | 69         |
| 1 to 2 hours           |  | 38%        | 48         |
| 3 hours or more        |  | 7%         | 9          |
| <b>Total Responses</b> |   |            | <b>126</b> |

**Figure 12. How often did the online discussion help finalize your submitted score?**

| Response                   | Chart   | Percentage | Count      |
|----------------------------|---|------------|------------|
| Never (0%)                 |  | 9%         | 11         |
| Rarely (up to 25%)         |  | 25%        | 31         |
| Occasionally (>25% to 50%) |  | 37%        | 47         |
| Often (> 50% to 75%)       |  | 13%        | 17         |
| Very Often (>75%)          |  | 13%        | 16         |
| Always (100%)              |  | 3%         | 4          |
| <b>Total Responses</b>     |   |            | <b>126</b> |

### Reviewing applications

While reviewing applications, reviewers spent most of their time reading the application (78 minutes) and less time writing their reviews (38 minutes). Contrary to grant programs, the Fellowship program has a structured review format and a much shorter grant application where reviewers are asked to provide brief comments on each criterion.

**Figure 13. How much time, on average, did you spend on:**

|   | Average in hours          | Standard Deviation |
|---|---------------------------|--------------------|
| a. Reading a single application                             | 1.30 hours (78 minutes)   | 1.12 (67 minutes)  |
| b. Writing the review of a single application               | 0.63 hours (37.8 minutes) | 0.43 (26 minutes)  |
| c. Reviewing a single application                           | 1.45 hours (87 minutes)   | 1.03 (62 minutes)  |
| d. Participating in an online discussion for an application | 0.42 hours (25 minutes)   | 0.37 (22 minutes)  |

### Peer Review Process & Technology

Most reviewers believed that the peer review process is efficient (84%) and that the technology is user-friendly (91%). However, numerous recommendations to improve the process and technology were raised and some will be implemented in future pilots. See the Summary of changes section above.

**Figure 14. Was the peer review process efficient?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Yes                    |  | 84%        | 118        |
| No                     |  | 16%        | 22         |
| <b>Total Responses</b> |   |            | <b>140</b> |

**Representative Comments: How can the peer review process be improved?**

- Cannot see scores, comments and application data when participating in the online discussion; must go back and forth between the screens;
- Provide options for some form of verbal communication;
- Encourage reviewers to submit reviews on time – or have consequences for not respecting the process;
- Standardized form of review – divide comments into categories: proposals, publications, etc.
- Common CV has excess irrelevant information. Awful format; some columns are too narrow, which makes reading entries cumbersome;
- Remote review process is fine for scholarships/fellowships at trainee stage but CANNOT possibly substitute for a face to face interaction which is necessary for grant applications;
- Reviewers should be evaluated;
- Shorter application.

**Figure 15. Was the technology employed to conduct the reviews user-friendly?**

| Response               | Chart   | Percentage | Count      |
|------------------------|---|------------|------------|
| Yes                    |  | 91%        | 126        |
| No                     |  | 9%         | 13         |
| <b>Total Responses</b> |   |            | <b>139</b> |

**Representative Comments: How can the technology employed to conduct the reviews be improved?**

- Ability to see comments/scores while completing the online discussion;
- CCV has duplicated information which appears in different sections with different formats. Presentation of the information complicates things;
- Scores should be updated as they are revised for other reviewers to see;
- Add video conferencing or give option to set up another form of communication.
- Asynchronous discussion is inefficient and unreliable;
- Designate a leader/moderator for the discussion.

## Appendix 2: Fellowship Reviewer Survey

Q 1. Would you classify yourself as a:

- Early career scientist (<5 years as an independent researcher)
- Mid-career scientist (5-10 years as an independent researcher)
- Senior scientist (>10 years as an independent researcher)

Q 2. Which research position(s) do you currently hold at this time?

Please select all that apply:

- Professor
- Associate Professor
- Assistant Professor
- Researcher
- Clinician
- Other (specify): \_\_\_\_\_

Q 3. Which of the following is your primary research domain:

- Biomedical
- Clinical
- Health systems/services
- Social, cultural, environmental and population health

Q 4 a) Did you look at the other reviews of the applications assigned to you?

- Yes
- No

Q 4 b) How much additional time in total (all applications) did you spend reading reviews of other reviewers?

- Under an hour
- 1 to 2 hours
- 3 hours or more

Q 4 c) How often did the reading of other reviews help finalize your submitted score?

- Never (0%)
- Rarely (up to 25%)
- Occasionally (>25% to 50%)
- Often (> 50% to 75%)
- Very Often (>75%)
- Always (100%)

Q 5 a) Did you initiate an online discussion thread?

- Yes
- No

Q 5 b) What factor(s) did you use to determine whether an online discussion was needed for an application?

- Choose all that apply
- Scoring discrepancy
- Content clarification required
- Quality check
- Other: \_\_\_\_\_

Q 6 a) Who should decide when an online discussion must take place?

- CIHR (e.g., using a decision rule within ResearchNet)
- Reviewers (at their discretion)
- Don't know

Q 6 b) What explicit criteria should be used to determine whether an online discussion must take place:

Q 7 a) Did you participate in an online discussion?

- Yes
- No

Q 7 b) Did online discussion increase the amount of time required to complete the review of an application (on average) in this competition?

- Under an hour
- 1 to 2 hours
- 3 hours or more

Q 7 c) How often did the online discussion help finalize your submitted score?

- Never (0%)
- Rarely (up to 25%)
- Occasionally (>25% to 50%)
- Often (> 50% to 75%)
- Very Often (>75%)
- Always (100%)

Q 8. How much time, on average, did you spend on (rounded to periods of 15 minutes):  
Please choose a value of "0" if the column is not applicable.

Hours | Minutes

- a) Reading a single application
- b) Writing the review of a single application
- c) Reviewing a single application
- d) Participating in an online discussion for an application

Q 9 a) Was the peer review process efficient?

- Yes
- No

Q 9 b) How can it be improved?

Q 10 a) Was the technology employed to conduct the reviews user-friendly?

- Yes
- No

Q 10 b) How can it be improved?

Q 11 a) Did you take the online peer review training for this Fellowships Pilot?

- Yes
- No

Q 11 b) How can it be improved?

Q 12 Do you have any additional feedback?