

1 **DORA-compliant measures to assess research quality and impact in biomedical institutions: review of**
2 **published research, international best practice and Delphi survey**

3
4 Measuring research activity, quality and impact

5
6 Anna R Gagliardi^{1*}, Rob HC Chen², Himani Boury,¹ Mathieu Albert³, James Chow⁴, Ralph S DaCosta⁵,
7 Michael Hoffman⁵, Behrang Keshavarz⁶, Pia Kontos⁶, Jenny Liu⁷, Mary Pat McAndrews⁷, Stephanie
8 Protze⁸

9
10 ¹ Toronto General Hospital Research Institute, University Health Network, Toronto, Ontario, Canada

11 ² UHN Research Solutions and Services, University Health Network, Toronto, Ontario, Canada

12 ³ The Institute for Education Research, University Health Network, Toronto, Ontario, Canada

13 ⁴ Techna Institute, University Health Network, Toronto, Ontario, Canada

14 ⁵ Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

15 ⁶ Toronto Rehabilitation Institute (KITE), University Health Network, Toronto, Ontario, Canada

16 ⁷ Krembil Research Institute, University Health Network, Toronto, Ontario, Canada

17 ⁸ McEwen Stem Cell Institute, University Health Network, Toronto, Ontario, Canada

18
19 * Corresponding author

20 Email: anna.gagliardi@uhnresearch.ca

21
22 **ABSTRACT**

23 **Objective**

24 The San Francisco Declaration on Research Assessment (DORA) advocates for assessing biomedical
25 research quality and impact, yet academic organizations continue to employ traditional measures such
26 as Journal Impact Factor. We aimed to identify and prioritize measures for assessing research quality
27 and impact.

28 **Methods**

29 We conducted a review of published and grey literature to identify measures of research quality and
30 impact, which we included in an online survey. We assembled a panel of researchers and research
31 leaders, and conducted a two-round Delphi survey to prioritize measures rated as high (rated 6 or 7 by ≥
32 80% of respondents) or moderate (rated 6 or 7 by ≥ 50% of respondents) importance.

33 **Results**

34 We identified 50 measures organized in 8 domains: relevance of the research program, challenges to
35 research program, or productivity, team/open science, funding, innovations, publications, other
36 dissemination, and impact. Rating of measures by 44 panelists (60%) in Round One and 24 (55%) in
37 Round Two of a Delphi survey resulted in consensus on the high importance of 5 measures: research
38 advances existing knowledge, research plan is innovative, an independent body of research (or
39 fundamental role) supported by peer-reviewed research funding, research outputs relevant to
40 discipline, and quality of the content of publications. Five measures achieved consensus on moderate
41 importance: challenges to research productivity, potential to improve health or healthcare, team
42 science, collaboration, and recognition by professional societies or academic bodies. There was high
43 congruence between researchers and research leaders across disciplines.

44 **Conclusions**

45 Our work contributes to the field by identifying 10 DORA-compliant measures of research quality and
46 impact, a more comprehensive and explicit set of measures than prior efforts. Research is needed to
47 identify strategies to overcome barriers of use of DORA-compliant measures, and to “de-implement”
48 traditional measures that do not uphold DORA principles yet are still in use.

49 INTRODUCTION

50 The San Francisco Declaration on Research Assessment (DORA) was established in 2012 during
51 the Annual Meeting of the American Society for Cell Biology [1]. DORA principles advocate for
52 reforming scientific research assessment based on a broad range of discipline-relevant
53 measures of quality and impact. A major tenet of DORA is the elimination of journal-based
54 metrics such as Journal Impact Factor. DORA recommends that academic organizations should
55 be explicit about criteria used for hiring, annual review, tenure, and promotion decisions;
56 assess the value and impact of all research outputs in addition to research publications; and
57 consider a broad range of measures including qualitative indicators of research impact such as
58 influence on policy and practice.

59
60 Research norms and outputs vary widely by discipline. Furthermore, journal publications
61 represent only one way to disseminate research, and metrics such as Journal Impact Factor are
62 skewed across disciplines. Therefore, reliance on such metrics is not an accurate,
63 comprehensive, or equitable way to judge the merits of a researcher, or research activity and
64 outputs. For example, qualitative research was rarely published in high-impact general medical
65 and health services research journals over a ten-year period [2]. Also, some journals are not
66 assigned an impact factor although they are peer-reviewed and listed in major research indices;
67 thus, reliance on Journal Impact Factor risks overlooking high-quality research published
68 outside of journals considered “high impact”. The harms associated with evaluating research
69 based on Journal Impact Factor and questions about the validity of how the impact score is
70 determined have long been recognized [3-8].

71 While the value of interdisciplinary or team science is widely recognized [9], research shows
72 that many researchers struggle to achieve legitimacy in biomedical settings [10]; for example,
73 qualitative health care researchers, whose political and epistemological orientation and
74 research processes are in opposition to positivism [11]. Some have noted that continued
75 reliance on the “latent biomedical conservatism that characterizes the health sciences”
76 combined with a lack of frameworks that acknowledge and properly assess diverse forms of
77 scholarship disadvantage many researchers and impede their professional advancement [12].
78
79 Given the many deficiencies of journal metrics for assessing research productivity and
80 contributions, it is no wonder that the DORA principles have been widely endorsed. As of March
81 23, 2022, 21,385 individuals and organizations in 156 countries are DORA signatories. However,
82 a challenge to implementing DORA principles is the lack of established alternatives to journal
83 metrics. For example, the Leiden Manifesto offers 10 principles that uphold DORA principles
84 (e.g. measure performance against research institute mission, account for variation by field in
85 publication and citation practices) [13]. A 2017 meeting of international experts in scientific
86 communication generated five principles upon which to judge research: assess contribution to
87 societal needs, employ responsible indicators, reward publishing of all research regardless of
88 the results, recognize the culture of open research, fund research that generates evidence on
89 optimal ways to assess science and faculty, and fund/recognize out-of-the-box ideas [14]. While
90 helpful in terms of guidance and advocacy, these principles may not represent a comprehensive
91 list of measures for assessing research quality and impact. Others have suggested criteria for
92 research assessment, but they are discipline-specific and not broadly applicable to diverse fields

93 of research. For example, Mazumdar et al. proposed measures to assess the contributions of
94 biostatisticians to team science [15].

95
96 As a DORA signatory, our organization formed a DORA advisory group (authors of this
97 manuscript) representing different research disciplines and stages of career to align research
98 reporting and assessment with DORA principles. To achieve this, we aimed to identify and
99 prioritize DORA-compliant measures for assessing research. We used an evidence- and
100 consensus-based approach that generated 10 measures and identified processes to support
101 uptake of those measures within our institution. Researchers in our organization and elsewhere
102 can employ these measures to describe and promote the value of their research, and academic
103 organizations can employ these measures and processes to support equitable hiring, annual
104 reviews, tenure, promotion and other decisions based on the quality and impact of research.
105 The purpose of this manuscript is to describe our methods and the resulting DORA-compliant
106 measures.

107

108 **METHODS**

109 **Research Design**

110 We conducted a sequential, multi-methods study. We assembled research assessment
111 measures and processes by conducting a scoping review of published and grey literature [16].
112 We chose a scoping review over other types of syntheses because it is characterized by the
113 inclusion of a range of study designs, which facilitates the exploration of literature in a given
114 field and reveals the nature of existing knowledge [17,18]. Similar in rigor to a systematic

115 review, a scoping review does not assess the methodological quality of included studies and
116 does not assume or generate a theoretical stance. We supplemented published research by
117 searching “grey” literature, referring to a range of types of documents (e.g. academic
118 publications, strategic plans, program evaluations) available on the Internet. Grey literature
119 searching is challenging because there are few dedicated repositories of grey literature, no
120 standard methods for searching for grey literature, and the effort required is inversely related
121 to the typically low yield [19,20]. However, we chose to do so for this study as the DORA
122 website provides links to reports of international initiatives that adopted and applied DORA
123 principles. Scoping review findings formed the basis of a Delphi survey. The Delphi technique is
124 a widely used method for generating consensus on strategies, recommendations, or quality
125 measures [21-23]. This technique is based on one or more rounds of survey in which panelists
126 independently rate recommendations until a degree of consensus is achieved. We did not
127 register a protocol. We consulted with the University Health Network Research Ethics Board,
128 who determined that we did not require ethics approval for this initiative. We complied with
129 research reporting criteria for scoping reviews [24] and Delphi studies [25].

130

131 **Scoping review**

132 *Eligibility*

133 Author ARG conducted a preliminary search in MEDLINE using the Medical Subject Headings
134 “employee performance appraisal” AND “research personnel” to become familiar with the
135 literature, draft eligibility criteria, and inform a more comprehensive search strategy. All
136 authors reviewed and refined PICO-based eligibility criteria. S1 Appendix details inclusion and

137 exclusion criteria. In brief, we included studies in which participants were researchers from a
138 wide array of research disciplines or research leaders based in academic settings. Research
139 disciplines reflected the *Canadian Research and Development Classification 2019* developed by
140 the Tri-Council Funding Agencies [26]. The issue referred to research productivity, contributions
141 to science, health systems or society, quality or impact, or other synonymous terms used by
142 eligible studies. Comparisons, or the purpose of assessment included hiring, annual review, re-
143 appointment, compensation, tenure, promotion, consideration for leadership or other awards,
144 etc. With respect to publication type, we included any qualitative, quantitative, or multiple-
145 /mixed-methods study. Outcomes included measures, indicators, criteria, suggestions,
146 recommendations, policies, or practices for research assessment; or the preferences of
147 researchers or research institutes for measures or processes related to research assessment.
148 We did not include measures reflecting the assessment of trainees or trainee research, non-
149 research measures (e.g. teaching, supervision, other services), or diversity due to a concurrent
150 effort underway at UHN with a focus on equity.

151

152 *Searching and screening*

153 ARG, who has medical librarian training, developed and executed searches (S2 Appendix),
154 complying with search strategy reporting guidelines [27]. We searched MEDLINE, EMBASE,
155 CINAHL, AMED, and the Web of Science for studies published in English language from 2013 to
156 January 15, 2021. We chose 2013 because DORA principles were published in 2012, following
157 which one might expect publications for research based on DORA principles. All search results
158 were imported into Covidence to remove duplicates and facilitate screening. Authors ARG, BK,

159 and MA independently screened the first 20 titles and abstracts and compared and discussed
160 findings. This identified only one discrepancy in the selection of eligible studies, which was
161 resolved by further refining eligibility criteria. Thereafter, ARG screened the remaining titles
162 and abstracts and acquired full-text versions of potentially-eligible articles.

163

164 *Data collection and analysis*

165 ARG extracted data from eligible articles on first author, year of publication, country, study
166 objective, research design and key findings (i.e. research assessment measures or related
167 processes). Based on that data, ARG compiled a list of unique measures and processes reported
168 or recommended across all included studies.

169

170 *Grey literature*

171 We employed a targeted approach to search for publicly available grey literature by browsing
172 the DORA website (<https://sfdora.org/>), following links from the DORA website to international
173 organizations, and both browsing and searching the websites of Canadian universities. On each
174 site, we searched for institutional policies, strategic plans, or other documents that described
175 research assessment measures, or reporting or evaluation processes. ARG searched for relevant
176 reports and extracted data on: organization name, title of the document or website, year
177 published, document purpose, and research assessment measures or processes. ARG compiled
178 a list of unique measures and processes, and integrated those with the list of measures and
179 processes compiled from published research, resulting in 49 unique measures organized in
180 eight categories that inductively emerged: relevance of research program, challenges to

181 productivity, team/open science, funding, innovations, publications, other dissemination, and
182 evidence of impact.

183

184 **Delphi survey**

185 *Survey development*

186 All authors reviewed the integrated list of measures, processes and refined wording. These
187 measures formed the basis of the Delphi survey that was administered using an in-house
188 application that creates online surveys.

189

190 *Sampling and recruitment*

191 A review of Delphi studies showed that the median number of panelists was 17 (range from 3
192 to 418) [23]. Other research found that the reliability of Delphi rating increased with panel size
193 [22]. To ensure that panelists represented multiple perspectives, we aimed to include persons
194 who varied by: research institute within our organization, research discipline, career stage
195 (early, mid, late), and professional role: researcher or research leader (e.g. head of the research
196 institute). To compile the list, we referred to institutional databases and asked research
197 institute administrators for suggestions. This resulted in a 74-member panel of researchers, of
198 which 6 (8.1%) were in leadership positions. Of the 74 panelists, 24 (32.4%) were early career,
199 28 (37.8%) midcareer and 22 (29.7%) late-career, referring to < 5 years, 5-10 years and > 10
200 years as independent researchers, respectively. Due to reasons of privacy and confidentiality,
201 we did not have access to data on gender, age or ethno-cultural characteristics of panelists.

202

203 *Data collection and analysis*

204 We asked panelists to rate each recommendation on a 7-point Likert scale (1=strongly disagree,
205 4=neutral, 7=strongly agree), comment on the relevance or wording of each recommendation if
206 desired, and suggest additional recommendations not included in the survey. Standard Delphi
207 protocol suggests that two rounds of rating with agreement by at least two-thirds of panelists
208 to either retain or discard items will prevent respondent fatigue and drop-out [20-22]. We
209 followed these suggestions and conducted two rounds of rating. We emailed Instructions and a
210 Round One survey link to panelists on July 21, 2021, with reminders at one and two weeks
211 following the initial invitation. Based on the results, we developed a Round One summary
212 report that included Likert scale response frequencies and comments for each
213 recommendation, which we organized by those retained (rated by at least 80% of panelists as 6
214 or 7), discarded (rated by at least 80% of panelists as 1 or 2) or no consensus (all others), along
215 with newly suggested recommendations. On September 15, 2021, we emailed panelists the
216 Round One summary report with a link to the Round Two survey, formatted similarly to the
217 Round One survey, to prompt rating of recommendations that did not achieve consensus for
218 inclusion or exclusion in Round One. We emailed a reminder at one, two and three weeks after
219 the initial invitation. We analyzed and summarized Round Two responses as described for
220 Round One. Ultimately, because few measures were highly rated by 80% of panelists, we
221 retained measures that achieved high (rated 6 or 7 by $\geq 80\%$ of panelists) or moderate (rated 6
222 or 7 by $\geq 50\%$ of panelists) consensus.

223

224 **RESULTS**

225 **Scoping review**

226 Of 1,566 unique search results, we excluded 1,538 titles, and of 28 potentially eligible full-text
227 articles, we excluded 17 due to a focus on publication metrics (9), ineligible publication type (6),
228 no research assessment measures or processes reported (1), and context was not biomedical
229 (1). Ultimately, 11 articles were included for review (Fig 1). S3 Appendix includes data extracted
230 from eligible articles [28-41]. Of 55 grey documents (27 from Canadian university websites; 28
231 from DORA and other international organizations), we excluded 29 because they did not
232 contain relevant content and included 26 documents in the review. S4 Appendix includes data
233 extracted from the eligible documents [42-62].

234

235 Fig 1

236

237 **Compiled measures and processes**

238 S5 Appendix shows the list of the 50 unique research assessment measures compiled from
239 published and grey literature. S6 Appendix shows the list of unique processes to support the
240 uptake of DORA-compliant measures. Table 1 includes select illustrative examples of those
241 processes.

242

243 Table 1. Select processes to support uptake of DORA-compliant measures

Category	Processes
Principles	<ul style="list-style-type: none">• Create a more porous research culture to promote interdisciplinary approaches, and enable more mobile and flexible research careers

	<ul style="list-style-type: none"> • Have individuals highlight and articulate their most meaningful contributions by describing the quality, significance and impact of their scholarship, and specify the level of impact: individual, community, system and population to help reviewers assess its merits
Responsibilities	<ul style="list-style-type: none"> • Research managers and administrators should champion these principles and the use of responsible evaluation within their institutions • Individual researchers are responsible for providing narrative descriptions of the quality, context, and impact of their research, individual research products, and other aspects of research without relying on surrogate metrics. The person best placed to articulate the importance of the research is the researcher themselves.
Review processes	<ul style="list-style-type: none"> • Assemble diverse review committees reflecting inclusivity, diversity, equity and ability—across gender, seniority, cultures, and under-represented minority populations—to bring a range of perspectives and experiences into decisions • To facilitate success, each Research Institute or Unit Chair shall: Review with each member their responsibilities and expectations; Meet with each member annually to discuss their annual report; and performance quality, progress and trajectory; Discuss career goals, and offer mentorship and other supports; Discuss merit

	recommendations and jointly agree on an action plan to address deficiencies
Implementation of DORA-compliant measures	<ul style="list-style-type: none"> • Provide education/training to research leaders and researchers about these principles, and how to assess (leaders) and describe (researchers) research productivity/contributions • Incentivize and reward a broader range of academic activities

244

245 **Delphi survey**

246 *Panelists*

247 Of 74 researchers invited to participate, 44 (59.5%) completed the Round One survey. Of those
 248 44, 24 (54.5%) completed the Round Two survey. Overall, 4 (66.7%) invited research leaders
 249 and another 40 (58.8%) invited researchers participated in at least one survey round (Table 2).

250

251 Table 2. Delphi respondent characteristics

Respondents	Survey Round	
	respondents/recipients (%)	
	One	Two
Role		
Researchers	40/68 (58.8)	22/40 (55.0)
Researchers in leadership role	4/6 (66.7)	2/4 (50.0)
Researchers		

Early career	11/24 (45.8)	5/11 (45.5)
Mid career	19/28 (67.9)	9/19 (47.4)
Late career	14/22 (63.6)	10/14 (71.4)
Total	44/74 (59.5)	24/44 (54.5)

252

253 *Delphi rating*

254 S7 Appendix shows respondent ratings of all measures and S8 Appendix lists the pros and cons
 255 offered by respondents for all measures that did not achieve high or moderate consensus.

256 Figure 2 summarizes results across two rounds of rating.

257

258 Fig 2

259

260 **Recommended measures**

261 Ten measures achieved high or moderate consensus (Table 3).

262

263 Table 3. Measures that achieved consensus to retain

Category	Measure (degree of consensus)
Relevance of research program	1. Research advances existing applied and/or theoretical knowledge (high)
	2. Research plan is innovative (e.g. generates novel methods, models, data, or other knowledge that addresses a noted gap) (high)
	3. Research directly addresses or has the potential to improve

	healthcare and the health of the public (moderate)
Funding	4. Evidence of research independence (e.g. PI or co-PI) OR of a fundamental role on a research team (e.g. biostatistician, qualitative researcher) with peer-reviewed research funding (high)
Innovations	5. Evidence of research outputs/products relevant to type of research (researcher can choose from the following list or provide other relevant options): (high) <ul style="list-style-type: none"> - Commercialization of technology (e.g. software, drugs, devices), launch of companies, invention disclosures, patent applications, issued patents - Creation of cohorts or registries - Reusable software or datasets - Reusable reagents (plasmids, mouse models, cell lines) - Clinical tests, algorithms or statistical models - Validated questionnaires or instruments - Contribution to policies, standards, guidelines or programs - Novel theory, model or framework - Novel research approaches, or methods - Other forms of achievement or outputs relevant to discipline (e.g. performance art)
Publications	6. Quality of the content of publications as judged by a peer reviewer or panel, in part based on rationale provided by the researcher of importance to their field (high)
Team Science	7. Where it would benefit the research topic or program, evidence of

	<p>local, regional, national or international team science (e.g. formation of clinical network or inter-/multi-disciplinary research group) (moderate)</p> <p>8. Evidence of collaboration or multidisciplinary research through participation as a co-investigator or mentor on other research teams (moderate)</p>
Recognition	9. Recognition by academic or professional societies (e.g. awards, honours) (moderate)
Challenges to research productivity	10. Qualitative description by researcher of challenges faced and mitigating strategies applied (e.g. leaves of absence for family or medical reasons) (moderate)

264

265 **DISCUSSION**

266 Academic organizations assess research activity and outputs, yet meaningful measures for
 267 doing so are lacking. In this study, we reviewed published research and grey literature to derive
 268 measures of research quality and impact that could be used to assess researcher activity and
 269 impact organized in 8 domains: relevance of research program, challenges to research program
 270 or productivity, team/open science, funding, innovations, publications, other dissemination,
 271 and impact. A two-round Delphi survey resulted in consensus on 10 measures, including 5
 272 measures of high importance: research advances existing knowledge, the research plan is
 273 innovative, an independent body of research (or fundamental role) supported by peer-
 274 reviewed research funding, research outputs relevant to the discipline, and quality of the

275 content of publications; and 5 measures of moderate importance: challenges to research
276 productivity, potential to improve health or healthcare, team science, collaboration, and
277 dissemination or recognition by professional societies or academic bodies. There was high
278 congruence between respondents with different roles (e.g. researcher, research leader)
279 sampled to represent a wide range of research disciplines.

280
281 Prior research that examined research assessment found that it relied on largely on journal
282 metrics and revealed few frameworks of assessment measures with little agreement across
283 frameworks. For example, a 2018 and survey of criteria used for assessing researchers at 92
284 international faculties of biomedical sciences revealed they largely employed traditional
285 measures such as the number of peer-reviewed publications, impact factor, and number or
286 amount of grant funding [22]. A 2018 scan of Canadian faculty of health sciences and medicine
287 websites identified few frameworks used to support hiring or promotion, and those identified
288 employed vague statements about creativity or quality but no explicit measures [12]. In a
289 survey of medicine and life sciences faculty at five Belgian universities, 126 respondents rated
290 publishing in high impact journals or publishing more papers than others as contributing more
291 to advancing careers rather than advancing science or personal satisfaction, and rated having
292 research results used or implemented higher on both scientific advancement and personal
293 satisfaction compared with career advancement [63]. A 2021 editorial on research impact
294 stated there are more than 20 frameworks to understand and evaluate research impact, but
295 noted they are context-specific, vary widely in the outcomes they emphasize and lack empirical

296 validation [64]. These studies underscore the lack of explicit non-metric based measures for
297 assessing research, a gap that our research addressed.

298

299 Our work builds on a 2017 meeting of 22 experts from the United States, England, Germany,
300 Netherlands and Canada who reviewed select literature critiquing traditional research
301 assessment and generated five principles upon which to judge research: societal benefit,
302 contributions to science, out-of-the-box ideas, full and transparent publication regardless of
303 results and open science [14]. Our work generated measures that match these first 3 principles
304 plus an additional 7 measures by which to assess research activity and outputs, as
305 recommended by DORA. Clearly, there is a paucity of research on non-traditional measures for
306 assessing research given that we identified only 11 empirical studies on this topic published
307 after the release of the DORA principles in 2012 [1]. Given a lack of insight on a range of
308 relevant measures for assessing research, our work contributes to the field by generating
309 consensus on non-traditional measures of research activity, quality, and impact that can be
310 used to uphold DORA principles in our organization and other academic organizations
311 worldwide who already endorsed DORA or are contemplating how to do so.

312

313 The 10 measures generated by this research can be used by researchers when reporting on the
314 quality and impact of their research, and by employers or evaluators when assessing
315 researchers for hiring, annual review, tenure, promotion and other decisions. Academic
316 research organizations and others (e.g. funders) can compare their research evaluation rubrics
317 and processes to the measures identified by this study as a means of planning or enhancing the

318 way that research is assessed. Of further support are the principles, responsibilities and
319 processes by which to apply these measures, and promote awareness, adoption and use of the
320 measures on the part of researchers and employers/evaluators. This work will be directly
321 relevant to the 21,385 individuals and organizations in 156 countries who have officially
322 endorsed DORA (as of March 23, 2022).

323

324 In a broader context, these findings are also germane to discussions about the tangible value of
325 research. Governments and funders worldwide are placing increasing emphasis on the
326 assessment of research impact to supply evidence of the value of their research investments to
327 society [65]. To foster research impact, national-level initiatives in the United Kingdom
328 (Collaborations for Leadership in Applied Health Research and Care) and the Netherlands
329 (Academic Collaborative Centres) invested heavily in implementing regional networks of
330 researchers or academic organizations, government policymakers, health system leaders, and
331 members of the public or representing healthcare advocacy groups [66,67]. These networks are
332 based on the concepts of participatory research or integrated knowledge translation, whereby
333 research is more likely to be relevant and used when planned from the outset with target users
334 [68]. Evaluations of these entities revealed they improved service delivery and associated
335 clinical outcomes [69]. In 2014, the United Kingdom established the Research Excellence
336 Framework, which defined research impact as: “an effect on, change or benefit to the
337 economy, society, culture, public policy or services, health, the environment or quality of life,
338 beyond academia.” [70]. The Framework was accompanied by over 6,000 case studies
339 demonstrating research impact. Analysis of a subset of high-impact case studies revealed the

340 most common forms of impact: practice (e.g. changing professional behaviour, and improving
341 organizational culture, quality of services, and outcomes), government policy (e.g. adopting
342 new policies, reducing costs), economic (e.g. greater revenue, profit or market share) and
343 public awareness (e.g. improving public knowledge or attention to an issue) [70]. UK Research
344 and Innovation is currently (as of December 2021) introducing a new Resume for Research and
345 Innovation for evaluating scientists that relies heavily on context instead of raw metrics (
346 <https://www.ukri.org/news/ukri-launches-new-resume-for-research-and-innovation/>). In addition to
347 initiatives like DORA [1], such national-level efforts that value research based on the good it can
348 achieve may well contribute to a declining reliance on journal metrics.

349
350 However, in this study, measures reflecting co-production of research with those outside of
351 academia; research reflecting the needs and preferences by sex, gender and intersectional
352 factors; and evidence of societal research impact were rated of low importance. Furthermore,
353 such practices are increasingly required by funders of health services research and related
354 disciplines, and if not included in assessment rubrics, may result in health services researchers
355 being held far more accountable than other disciplines for resource- and time-intensive
356 activities that go unrecognized by employers or assessors who continue to rely on traditional
357 research metrics. The adoption of new practices can be slow, particularly when the necessary
358 change requires a profound culture shift, as is the case with DORA principles of research
359 assessment. Thus, additional research is needed to understand the perceived and actual value
360 of DORA principles and our measures, barriers to their uptake, and the knowledge and
361 strategies needed to address these barriers. This will be critical to informing interventions that

362 support the embracing of new measures of research quality and impact, and to “de-implement”
363 traditional measures that are inconsistent with DORA and deemed inappropriate, yet are still in
364 use. One way to do this is to learn how other organizations who have successfully adopted
365 DORA-compliant measures and processes achieved the culture shift. Recognizing that culture
366 shift may be a major barrier to adopting the measures recommended in this report, ongoing
367 research is needed to assess the perceived value of these measures and barriers to their
368 uptake, knowledge needed to select and tailor strategies or interventions aimed at supporting
369 uptake. For example, measures must be reported qualitatively by researchers, and judged
370 qualitatively by those with expertise in a relevant discipline, which can be more involved and
371 time-consuming than traditional quantitative metrics such as counting number of publications.
372 Because changes in research assessment may have broader implications, interviews should also
373 be conducted with non-researcher staff such as human resources, or managers responsible for
374 compiling and analyzing annual or periodic research activity reports submitted by researchers.
375 Also, forging strategic and tactical alliances with academic organizations, publishers, and
376 funding bodies will be necessary to achieve the successful uptake of non-traditional measures.
377
378 This study features several strengths. The measures rated by panelists were derived from
379 research and international best practices. We assembled a panel comprised of researchers
380 representing different research roles and disciplines. The large panel size enhanced reliability.
381 Two rounds of rating minimized respondent fatigue, which achieved a high response rate in
382 both rounds. We optimized rigor by complying with methodology and reporting criteria for
383 scoping reviews and Delphi studies [24,25]. Findings are bolstered by the high congruence in

384 rating between researchers, research leaders, and those representing different research
385 disciplines and career stages. We must also acknowledge some limitations. Our search for
386 sources of measures may not have been sufficiently comprehensive and only 11 papers on the
387 subject published since 2012 emerged; however, as part of the Delphi process, panelists were
388 asked to identify additional measures not already included in the survey. For reasons of privacy
389 and confidentiality, we did not have access to respondents' personal details, and therefore
390 could not examine ratings of measures by gender, age or ethno-cultural characteristics;
391 however, ratings were congruent, so sub-analyses may not have yielded meaningful
392 differences. Respondents' views may differ from those of researchers or research leaders in
393 other jurisdictions. The findings may not be generalizable in countries outside of Canada with
394 differing scientific or academic cultures and structures. However, numerous organizations
395 worldwide have embraced DORA, so the measures generated in our work are likely relevant at
396 organizational level.

397
398 In conclusion, a two-round Delphi survey of researchers and research leaders representing a
399 range of scientific disciplines, based on compilation of measures of research assessment from
400 published and grey literature, resulted in consensus on ten measures compliant with DORA
401 principles that can be used by researchers to report on the quality and impact of their research
402 activity, and by employers/evaluators to assess researchers for hiring, annual review and
403 promotion or tenure decisions.

404

405 **ACKNOWLEDGEMENTS**

406 We acknowledge Joan Wither and Catriona Steele, who contributed to early-stage decision-
407 making, and Stephanie Susman, who assisted with data collection.

408

409 REFERENCES

- 410 1. San Francisco Declaration on Research Assessment [Internet]. DORA. 2012 [cited 2022 Jun
411 2]. Available from: <https://sfdora.org/>
- 412 2. Gagliardi AR, Dobrow MJ. Paucity of qualitative research in general medical and health
413 services and policy research journals: analysis of publication rates. *BMC Health Serv Res.*
414 2011;11:268.
- 415 3. Gingras Y. *Bibliometrics and Research Evaluation: Uses and abuses*. Cambridge, MA: The
416 MIT Press; 1992.
- 417 4. Muller JZ. *The tyranny of metrics*. Princeton, New Jersey: Princeton University Press; 2019
- 418 5. Seglen PO. Why the impact factor of journals should not be used for evaluating
419 research. *BMJ.* 1997;314:498-502.
- 420 6. Not-so-deep impact. *Nature.* 2005;435:1003-1004.
- 421 7. The PLOS Medicine Editors. The Impact Factor Game. *PLOS Med.* 2006;3:6.
- 422 8. Rossner M, Van Epps H, Hill E. Show me the Data. *J Cell Biol.* 2007;179(6):1091-1092.
- 423 9. Hall KL, Vogel AL, Huang GC, Serrano KJ, Rice EL, Tsakraklides SP, et al. The Science of Team
424 Science: A Review of the Empirical Evidence and Research Gaps on Collaboration in Science.
425 *Am Psychol.* 2018;73:532-548.

- 426 10. Albert M, Paradis E, Kuper A. Interdisciplinary promises versus practices in medicine: The
427 decoupled experiences of social sciences and humanities scholars. *Soc Sci Med*. 2015;126:
428 17-25.
- 429 11. Kontos P, Grigorovich A. “Sleight of Hand” or “Selling Our Soul”? Surviving and Thriving as
430 Critical Qualitative Health Researchers in a Positivist World. *FQS* [Internet]. 2018 May 25
431 [cited 2022 Jun 2]. Available from: <https://www.qualitative>
432 [research.net/index.php/fqs/article/view/2990](https://www.qualitative-research.net/index.php/fqs/article/view/2990)
- 433 12. Webster F, Gastaldo D, Durant S, Eakin J, Gladstone B, Parsons J, et al. Doing Science
434 Differently: A Framework for Assessing the Careers of Qualitative Scholars in the Health
435 Sciences. *Int J Qual Methods*. 2019;18:1-7.
- 436 13. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for
437 research metrics. *Nature*. 2015;520:429-431.
- 438 14. Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JP, Goodman SN. Assessing scientists
439 for hiring, promotion, and tenure. *PLOS Biol*. 2018;16:3.
- 440 15. Mazumdar M, Messinger S, Finkelstein DM, et al. Evaluating Academic Scientists
441 Collaborating in Team-Based Research: A Proposed Framework. *Acad Med*. 2015;90:1302-
442 1308.
- 443 16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International J*
444 *Soc Res Methodol*. 2005;8:19-32.
- 445 17. O'Brien KK, Colquhoun H, Levac D, Baxter L, Tricco AC, Straus S, et al. Advancing scoping
446 study methodology: a web-based survey and consultation of perceptions on terminology,
447 definition and methodological steps. *BMC Health Serv Res*. 2016;16:305.

- 448 18. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or
449 scoping review? Guidance for authors when choosing between a systematic or scoping
450 review approach. *BMC Med Res Methodol.* 2018;18:143.
- 451 19. Benzies KM, Premji S, Hayden KA, Serrett K. State-of-the-evidence reviews: advantages and
452 challenges of including grey literature. *Worldviews Evid Based Nurs.* 2006;3:55-61.
- 453 20. Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al. Searching
454 and synthesising 'grey literature' and 'grey information' in public health: critical reflections
455 on three case studies. *Syst Rev.* 2016;5:164.
- 456 21. Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services
457 research. *BMJ.*1995;311:376-380.
- 458 22. Boukdedid R, Abdoul H, Loustau M, et al. Using and reporting the Delphi method for
459 selecting healthcare quality indicators: a systematic review. *PLOS One.* 2011;6: 1-9.
- 460 23. Stelfox HT, Straus SE. Measuring quality of care: considering conceptual approaches to
461 quality indicator development and evaluation. *J Clin Epidemiol.* 2013;66:1328-1337.
- 462 24. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR):
463 Checklist and Explanation. *Ann Intern Med.* 2018;169:467-473.
- 464 25. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting
465 DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological
466 systematic review. *Palliat Med.* 2017;31:684-706.
- 467 26. Statistics Canada. Canadian research and Development Classification (CRDC) 2020 version
468 1.0 [Internet]. Statistics Canada. 2020 [cited 2022 Jun 6]. Available from:
469 <https://www.statcan.gc.ca/en/subjects/standard/crdc/2020v1/index>

- 470 27. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review
471 of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol.* 2016;75:40-46.
- 472 28. Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, et al.
473 Documenting Social Media Engagement as Scholarship: A New Model for Assessing
474 Academic Accomplishment for the Health Professions. *J Med Internet Res.* 2020;22: e25070.
- 475 29. Clement L, Dorman JB, McGee R. The Academic Career Readiness Assessment: Clarifying
476 Hiring and training expectations for future Biomedical Life Sciences Faculty. *CBE—Life Sci*
477 *Educ.* 2020;19:1-22.
- 478 30. Husain A, Repanshek Z, Singh M, et al. Consensus Guidelines for Digital Scholarship in
479 Academic Promotion. *West J Emerg Med.* 2020;21:883-891.
- 480 31. Rice D, Raffoul H, Ioannidis J, Moher D. Academic criteria for promotion and tenure in
481 biomedical sciences faculties: cross sectional analysis of international sample of universities.
482 *BMJ.* 2020;369:m2081.
- 483 32. Brody AA, Bryant AL, Perez GA, Bailey DE. Best practices and inclusion of team science
484 principles in appointment promotion and tenure documents in research intensive schools of
485 nursing. *Nurs Outlook.* 2019;67:133-139.
- 486 33. Klein KC, Kelling SE, Pais K, Lee CA, Bostwick JR. From clinical assistant to clinical associate
487 professor: Examination of a sample of promotion guidelines. *Curr Pharm Teach Learn.*
488 2019;11:346-351.
- 489 34. LeMaire SA, Trautner BW, Ramamurthy U, et al. An Academic Relative Value Unit System for
490 Incentivizing the Academic Productivity of Surgery Faculty Members. *Ann Surg.*
491 2018;268:526-533.

- 492 35. Sehgal NL, Neeman N, King TE. Early Experiences After Adopting a Quality Improvement
493 Portfolio Into the Academic Advancement Process. *Acad Med*. 2017;92:78-82.
- 494 36. Finney JW, Amundson EO, Bi X, et al. Evaluating the Productivity of VA, NIH, and AHRQ
495 Health Services Research Career Development Awardees. *Acad Med*. 2016;91:563-569
- 496 37. Kairouz VF, Raad D, Fudyma J, Curtis AB, Schünemann HJ, Akl EA. Assessment of faculty
497 productivity in academic departments of medicine in the United States: a national survey.
498 *BMC Med Educ*. 2014;14:205.
- 499 38. University of Calgary. [Internet]. Criteria For Appointment, Promotion, Merit Increment and
500 Tenure of Full-time Faculty. University of Calgary. 2008 Dec [cited 2022 Jun 6]. Available
501 from:
502 https://www.ucalgary.ca/hr/sites/default/files/teams/239/medicine_faculty_guidelines.pdf
- 503 39. University of Victoria. University Of Victoria Faculty of Human and Social Development
504 Faculty Evaluation Policy 2019 – 2022 [Internet]. 2020 [cited 2022 Jun 7]. Available from:
505 [https://www.uvic.ca/hsd/assets/docs/policies/approvedhsd20192022facultyevaluationpolic](https://www.uvic.ca/hsd/assets/docs/policies/approvedhsd20192022facultyevaluationpolicy16jan2020.pdf)
506 [y16jan2020.pdf](https://www.uvic.ca/hsd/assets/docs/policies/approvedhsd20192022facultyevaluationpolicy16jan2020.pdf)
- 507 40. University of Regina. Criteria document for faculty members and instructors-Terms of
508 reference for assignment of duties, performance review, career progress, and sabbaticals
509 [Internet]. 2020 Sep 29 [cited 2022 Jun 7]. Available from:
510 [https://www.uregina.ca/science/assets/docs/pdf/2017-Criteria-Document-FacultyInstr-](https://www.uregina.ca/science/assets/docs/pdf/2017-Criteria-Document-FacultyInstr-29VIII2017%20DRAFT.pdf)
511 [29VIII2017%20DRAFT.pdf](https://www.uregina.ca/science/assets/docs/pdf/2017-Criteria-Document-FacultyInstr-29VIII2017%20DRAFT.pdf)
- 512 41. University of Alberta. Procedures and criteria for tenure, promotion, merit and sabbaticals
513 [Internet]. 2017 Nov 29 [cited 2022 Jun 7]. Available from:

- 514 <https://www.ualberta.ca/pediatrics/media-library/people/faculty-development/tenure->
515 [track-promotions/fomd-fec-standards-2017-approved.pdf](https://www.ualberta.ca/pediatrics/media-library/people/faculty-development/tenure-track-promotions/fomd-fec-standards-2017-approved.pdf)
- 516 42. Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S, et al. The Metric Tide: Report of the
517 Independent Review of the Role of Metrics in Research Assessment and Management.
518 2015.
- 519 43. Metrics toolkit. [cited 2022 Jun 8]. Available from: <https://www.metrics-toolkit.org/>
- 520 44. Stacey P. International Network of Research Management Societies [Internet]. INORMS.
521 INORMS; 2022 [cited 2022 Jun 8]. Available from: <https://inorms.net/>
- 522 45. Research Quality Plus. A Holistic Approach to Evaluating Research [Internet]. 2016 [cited
523 2022 Jun 9]. Available from: <https://sfdora.org/wp-content/uploads/2020/11/IDL-56528.pdf>
- 524 46. The Royal Society. Research culture: changing expectations. 2019 April [cited 2022 Jun 8].
525 Available from: <https://royalsociety.org/-/media/policy/projects/changing->
526 [expectations/changing-expectations-conference-report.pdf](https://royalsociety.org/-/media/policy/projects/changing-expectations/changing-expectations-conference-report.pdf)
- 527 47. Canadian Academy of Health Sciences. Making an impact - A Preferred Framework and
528 Indicators to Measure Returns on Investment in Health Research [Internet]. Canadian
529 Academy of Health Sciences. 2009 [cited 2022 Jun 8]. Available from: <https://www.caahs->
530 [acss.ca/wp-content/uploads/2011/09/ROI_FullReport.pdf](https://www.caahs-acss.ca/wp-content/uploads/2011/09/ROI_FullReport.pdf)
- 531 48. Saenen B, Hatch A, Curry S, Proudman V, Lakoduk A. Case Study Report-Reimagining
532 Academic Career Assessment: Stories of innovation and change [Internet]. 2021 [cited 2022
533 Jun 8]. Available from: <https://eua.eu/downloads/publications/eua-dora->
534 [sparc_case%20study%20report.pdf](https://eua.eu/downloads/publications/eua-dora-sparc_case%20study%20report.pdf)

- 535 49. Universitat Oberta de Catalunya. Open knowledge action plan: Frame of action - UOC
536 [Internet]. Universitat Oberta de Catalunya. 2019 [cited 2022 Jun 8]. Available from:
537 [https://www.uoc.edu/portal/resources/EN/documents/coneixement-obert/pla-accio-](https://www.uoc.edu/portal/resources/EN/documents/coneixement-obert/pla-accio-coneixement-obert.pdf)
538 [coneixement-obert.pdf](https://www.uoc.edu/portal/resources/EN/documents/coneixement-obert/pla-accio-coneixement-obert.pdf)
- 539 50. Universiteit. Vision Statement for Evaluating Research at Ghent University [Internet].
540 Universiteit Gent. 2016 [cited 2022 Jun 8]. Available from:
541 <https://www.ugent.be/en/research/research-strategy/evaluation>
- 542 51. Good Practice in Researcher Evaluation. Recommendation For the Responsible Evaluation of
543 a Researcher in Finland [Internet]. Responsible Research Series. 2020 [cited 2022 Jun 8].
544 Available from: <https://avointiede.fi/sites/default/files/2020-03/responsible-evaluation.pdf>
545 Universities Norway Consortium, Norway.
- 546 52. The Working Group on Rewards under Open Source. Evaluation of research careers fully
547 acknowledging open science practices: Rewards, incentives and/or recognition for
548 researchers practicing open science. [Internet]. Photo of Publications Office of the European
549 Union. Publications Office of the European Union; 2017 [cited 2022 Jun 8]. Available from:
550 [https://op.europa.eu/en/publication-detail/-/publication/47a3a330-c9cb-11e7-8e69-](https://op.europa.eu/en/publication-detail/-/publication/47a3a330-c9cb-11e7-8e69-01aa75ed71a1/language-en)
551 [01aa75ed71a1/language-en](https://op.europa.eu/en/publication-detail/-/publication/47a3a330-c9cb-11e7-8e69-01aa75ed71a1/language-en)
- 552 53. University College London. UCL Academic Careers Framework [Internet]. 2018 [cited 2022
553 Jun 11]. Available from: <https://www.ucl.ac.uk/human-resources/>
- 554 54. UMC Utrecht. Guide for reviewers/evaluators that use the UMC Utrecht indicators for
555 impact [Internet]. UMC Utrecht. 2016 [cited 2022 Jun 8]. Available from: <https://assets-eu->

- 556 [01.kc-usercontent.com/546dd520-97db-01b7-154d-79bb6d950a2d/a2704152-2d16-4f40-](https://www.kc-usercontent.com/546dd520-97db-01b7-154d-79bb6d950a2d/a2704152-2d16-4f40-9a4b-33db23d1353e/Format-Impact-indicator-evaluation-pilot-incl-introduction.pdf)
- 557 [9a4b-33db23d1353e/Format-Impact-indicator-evaluation-pilot-incl-introduction.pdf](https://www.kc-usercontent.com/546dd520-97db-01b7-154d-79bb6d950a2d/a2704152-2d16-4f40-9a4b-33db23d1353e/Format-Impact-indicator-evaluation-pilot-incl-introduction.pdf)
- 558 55. University of Bath. Principles of Research Assessment and Management [Internet].
- 559 University of Bath. 2021 [cited 2022 Jun 8]. Available from:
- 560 [https://www.bath.ac.uk/corporate-information/principles-of-research-assessment-and-](https://www.bath.ac.uk/corporate-information/principles-of-research-assessment-and-management/)
- 561 [management/](https://www.bath.ac.uk/corporate-information/principles-of-research-assessment-and-management/)
- 562 56. UCI Academic Personnel. Identifying faculty contributions to collaborative scholarship
- 563 [Internet]. UCI Academic Personnel. 2019 [cited 2022 Jun 8]. Available from:
- 564 <https://ap.uci.edu/faculty/guidance/collaborativescholarship/>
- 565 57. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US Biomedical Research from its
- 566 systemic flaws. *Proceedings of the National Academy of Sciences*. 2014;111(16):5773–5777.
- 567 58. FWF. Application Guidelines for Stand-Alone Projects (valid from 15 March 2022) [Internet].
- 568 FWF Der Wissenschaftsfonds. 2022 [cited 2022 Jun 8]. Available from:
- 569 https://www.fwf.ac.at/fileadmin/files/Dokumente/Antragstellung/Einzelprojekte/p_applica
- 570 [tion-guidelines.pdf](https://www.fwf.ac.at/fileadmin/files/Dokumente/Antragstellung/Einzelprojekte/p_applica)
- 571 59. NHMRC. Guide to NHMRC PEER review 2018 [Internet]. Building a Healthy Australia. 2018
- 572 [cited 2022Jun8]. Available from:
- 573 [https://www.nhmrc.gov.au/sites/default/files/documents/attachments/guide-nhmrc-peer-](https://www.nhmrc.gov.au/sites/default/files/documents/attachments/guide-nhmrc-peer-review-2018.pdf)
- 574 [review-2018.pdf](https://www.nhmrc.gov.au/sites/default/files/documents/attachments/guide-nhmrc-peer-review-2018.pdf)
- 575 60. VSNU, KNAW, NWO. Strategy evaluation protocol - Vereniging van Universiteiten [Internet].
- 576 Universiteiten van Nederland. 2020 [cited 2022Jun11]. Available from:

- 577 <https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/SEP>
- 578 [2021-2027.pdf](#)
- 579 61. European Commission, Directorate-General for Research and Innovation. Evaluation of
580 research careers fully acknowledging Open Science practices : rewards, incentives and/or
581 recognition for researchers practicing Open Science. Publications Office; 2017.
- 582 62. Wellcome Trust. Open access policy 2020 - wellcome [Internet]. 2018 [cited 2022 Jun 7].
583 Available from: [https://wellcome.org/sites/default/files/wellcome-open-access-policy-](https://wellcome.org/sites/default/files/wellcome-open-access-policy-2020.pdf)
584 [2020.pdf](#)
- 585 63. Aubert Bonn N, Pinxten W. Advancing science or advancing careers? researchers' opinions
586 on Success Indicators. PLOS One. 2021;16:2.
- 587 64. Büttner F, Ardern CL, Blazey P, et al. Counting publications and citations is not just
588 irrelevant: it is an incentive that subverts the impact of clinical research. Br J Sports Med.
589 2021;55:647-648.
- 590 65. OECD. Reference framework for assessing the scientific and socio-economic impact of
591 research infrastructures (Internet). 2019 [cited 2022 June 12]. Available from:
592 [https://www.oecd-ilibrary.org/science-and-technology/reference-framework-for-assessing-](https://www.oecd-ilibrary.org/science-and-technology/reference-framework-for-assessing-the-scientific-and-socio-economic-impact-of-research-infrastructures_3ffee43b-en)
593 [the-scientific-and-socio-economic-impact-of-research-infrastructures_3ffee43b-en](#)
- 594 66. Martin GP, McNicol S, Chew S. Towards a new paradigm in health research and practice?
595 Collaborations for Leadership in Applied Health Research and Care. J Health Organ Manag.
596 2013;27:193-208.

- 597 67. Hoeijmakers M, Harting J, Jansen M. Academic Collaborative Centre limburg: A platform for
598 knowledge transfer and exchange in Public Health Policy, research and practice? Health
599 Policy. 2013;111:175-183.
- 600 68. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation
601 (IKT) in health care: a scoping review. Implementation Science. 2016;11(1):38. Oborn E,
602 Barrett M, Prince K, Racko G. Balancing exploration and exploitation in transferring research
603 into practice: a comparison of five knowledge translation entity archetypes. Implement Sci.
604 2013;8:104.
- 605 69. Jensen EA, Wong P, Reed MS. How research data deliver non-academic impacts: A
606 secondary analysis of UK Research Excellence Framework Impact Case Studies. PLOS One.
607 2022;17:3.

608

609 **SUPPORTING INFORMATION CAPTIONS**

610 Fig 1. PRISMA diagram of articles included in review of published research

611 Fig 2. Delphi process and results

612 S1 Appendix. Eligibility criteria for published research

613 S2 Appendix. Search strategy for published research

614 S3 Appendix. Data extracted from included articles

615 S4 Appendix. Eligible documents identified in grey literature

616 S5 Appendix. Research assessment measures compiled from published and grey literature

617 S6 Appendix. Processes to support uptake of research assessment measures

618 S7 Appendix. Respondent ratings of all measures

619 S8 Appendix. Pros and cons reported by respondents for measures not prioritized