

Formative, embedded evaluation to strengthen interdisciplinary team science: Results of a 4-year, mixed methods, multi-country case study

Susan Roelofs^{1,*}, Nancy Edwards¹, Sarah Viehbeck^{2,3}
and Cody Anderson⁴

¹School of Nursing, University of Ottawa, 1 Stewart St., Room 212, Ottawa, ON K1N 6N5, Canada, ²School of Public Health and Health Systems, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada, ³Interdisciplinary School of Health Sciences, University of Ottawa, 1 Stewart St., Room 212, Ottawa, ON K1N 6N5, Canada, and ⁴Research Advisor, Public Safety Canada, 269 Laurier Avenue West, Ottawa, ON K1A 0P8, Canada

*Corresponding author: Email: sroelofs@uottawa.ca

Abstract

Evaluation of interdisciplinary, team science research initiatives is an evolving and challenging field. This descriptive, longitudinal, mixed methods case study examined how an embedded, formative evaluation approach contributed to team science in the interdisciplinary Research into Policy to Enhance Physical Activity (REPOPA) project, which focused on physical activity policy-making in six European countries with divergent policy systems and researcher–policymaker networks. We assessed internal project collaboration, communication, and networking in four annual data collection cycles with REPOPA team members. Data were collected using work package team and individual interviews, and quantitative collaboration and social network questionnaires. Interviews were content analyzed; social networks among team members and with external stakeholder were examined; collaboration scores were compared across 4 years using analysis of variance (ANOVA). Annual monitoring reports with action recommendations were prepared and discussed with consortium members. Results revealed consistently high response rates. Collaboration and communication scores, high at baseline, improved slightly, but ANOVA results were nonsignificant. Internal network changes tracked closely with implementation progress. External stakeholders were primarily governmental, with a marked shift from local/provincial level to national/international during the project. Diversity (disciplinary, organizational, and geopolitical) was a project asset influencing and also challenging collaboration, implementation, and knowledge translation strategies. In conclusion, formative evaluation using an embedded, participatory approach demonstrated utility, acceptability, and researcher engagement. A trusting relationship between evaluators and other project members built on joint identification of team science objectives for the evaluation at project outset, codeveloping guiding principles, and encouraging team reflexivity throughout the evaluation.

Key words: formative evaluation; embedded evaluation; team science; interdisciplinary research; collaboration; diversity.

1. Introduction

Interdisciplinary team science is regarded as essential for tackling the complex, multifaceted problems investigated by wide-ranging areas of research. Team science is collaborative research involving a

purposeful intersection of disciplines, methods, and often sectors (National Research Council 2015). By leveraging a multiplicity of scientific backgrounds and expertise, this approach aims to strengthen innovations in research approaches and outputs (Bennett and

Gadlin 2012; Baker 2015). However, the science of team science has been characterized as nascent (Lotrecchiano 2013), and evaluation approaches continue to evolve (Falk-Krzesinski et al. 2010; de Jong et al. 2011; Belcher et al. 2016).

Evaluating team science initiatives is challenging, due not only to inherent complexities of interdisciplinary endeavors but also to the range of settings where research is done and the mix of organizations where researchers are employed (Klein 2008; Trochim et al. 2008). There have been notable advances in assessing and understanding the benefits and challenges of collaborative research (Bennett, Gadlin and Levine-Finley 2010), how team science can advance innovation, what scientific and policy outcomes are achieved, and how these are influenced by proximity (geographic, cognitive, social, organizational, and institutional) (Boschma 2005). Team processes can influence project results (Prokopy et al. 2015), although an examination of these processes has been outside the mandate of most summative impact evaluations (National Research Council 2015). Furthermore, the predominant, cross-sectional orientation of these impact evaluations is not conducive to examining shifts in team processes that occur during multiyear projects.

Ongoing course corrections are often needed to strengthen collaborative approaches during research implementation. Formative evaluation, particularly when grounded in a participatory approach, may enhance project effectiveness (Stetler et al. 2006; Cousins, Whitmore and Shulha 2013). However, the contribution of formative evaluation processes to learning about good team science has seldom been described, and we could find no published studies describing embedded evaluators in these formative studies.

This article uses the case of a 4-year, multi-country, interdisciplinary research project, distinctive for its interventions in divergent, real-life policymaking contexts. The primary objectives of this article are to describe the acceptability and utility of our formative and embedded evaluation approach, and to examine whether and how formative results and related recommendations strengthened collaboration among team members. We used a longitudinal case study design with mixed methods data collection to address these objectives.

The article is structured as follows. We begin with an examination of the literature describing influences on, and evaluation approaches for, team science initiatives. The second section describes the team science case study (the Research into Policy to Enhance Physical Activity (REPOPA) project) and its formative and embedded evaluation strategy. This is followed by the longitudinal case study methods, results, and discussion.

2. Literature review

2.1 Influences on team science

Effective team science requires deliberately building in and managing functional diversity, those variations among team members' backgrounds, skills, and abilities (Cheung et al. 2016). Although disciplinary diversity is an essential attribute of team science collaborations especially for basic and applied research, team members may also differ in nationality, career stage, gender, and collaborative experience (Hall et al. 2008; Cheruvelil et al. 2014).

Fields such as economic geography examined how spatial and nonspatial dimensions of diversity influence collaborative innovation and learning, connections, and network involvement (Heringa et al. 2014; Davids and Frenken 2018). Boschma (2005) set out five

proximity dimensions: geographic (physical distance), cognitive (shared explicit and tacit knowledge), social (trust and cooperation), organizational (organizational structure), and institutional (cultural, legal, and incentive frameworks). Questions about the relative importance of different forms of proximity to collaborative learning and innovation, and what constitutes optimal proximity, are still being explored. While some degree of proximity is needed for connection and network engagement, the 'proximity paradox' suggests that after a certain point, too much proximity inhibits innovation through lock-in. Learning and novel insights require some degree of distance (i.e. differences) among those collaborating (Davids and Frenken 2018).

For team science projects, which involve long-distance collaborations, geographic proximity may be important for establishing connections, but combinations of close proximity on other dimensions (e.g. high levels of trust and cooperation) can play a stronger supporting role. This may be especially the case for 'mode 2' projects: interdisciplinary efforts aimed at solving real-world problems where utilization of research knowledge is an essential aspect, strongly influenced by a dynamic context (in contrast to 'mode 1' basic scientific discovery) (Gibbons 2000). Hardeman et al. (2015) found geographic proximity was particularly important for 'mode 1' knowledge production, while other forms of proximity mattered for 'mode 2' knowledge production. In a review of sociological research on the formation, persistence, and dissolution of social networks, Rivera, Soderstrom, and Uzzi (2010) identified geographic proximity and two other mechanisms influencing network activity relevant to the sharing and creation of knowledge: relational mechanisms (direct and indirect connections among individuals) and assertive mechanisms (divergence and similarity among individuals' attributes).

That functionally diverse teams will realize their innovative potential is not a given. The challenge of interdisciplinary research lies in how to realize the latent potential of diverse teams to go beyond the separate but cooperative disciplinary contributions that mark multidisciplinary initiatives to achieving knowledge integration, synthesis, and innovation (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine 2005; Siedlok and Hibbert 2014). There is still much to learn about how to accomplish this. A National Academies report into the state of team science (National Research Council 2015) drew on research from the social sciences and basic and applied science to identify seven key challenges to team effectiveness: diverse members, deep knowledge integration, large size, goal misalignment, permeability of membership boundaries, geographic dispersal, and high task interdependence.

Organizational psychology studies have identified moderators of team innovation such as affect-based trust and transformational leadership; these exert their influence by strengthening knowledge-sharing and reducing conflict within teams (Hüttermann and Boerner 2011; Reuveni and Vashdi 2015). Task network structures (how work is structured across a team) also shape team potency and performance, an effect that is heightened when teams are culturally diverse (Tröster, Mehra and van Knippenberg 2014); however, the question of what comprises optimum task network density is still being explored.

Team reflexivity, 'the extent to which teams collectively reflect upon and adapt their working methods and functioning' (Schippers, West and Dawson 2015), has been identified as an important

predictor of innovation and can improve team learning, decision-making, outcomes, and team effectiveness (Konradt et al. 2016). Achieving high-performing collaborative research teams (Cheruvilil et al. 2014) requires attention to collaborative, communication, and interpersonal processes that affect how diverse teams work together to influence knowledge integration and research outcomes (Klein 2008).

Team challenges may be compounded in projects that involve multiple study sites, span different country contexts, and encompass organizational cultures with variable role expectations for scientists (Cheruvilil et al. 2014; Goring et al. 2014). Developing collaborative processes takes time; members need a commitment to the ‘slogging through’ phase to develop a common cross-disciplinary language, a willingness to use research methods that may be outside individual scientific comfort zones, and specific strategies to foster collaboration among investigators at different centers (Okamoto 2015; Klink et al. 2017). Furthermore, addressing these challenges may be costly. Cummings and Kiesler (2007) found that projects involving researchers from multiple institutions bore higher coordination costs for task division, resource sharing, knowledge production, and communication, related to institutional differences and proximity. Others have found that as team size grows, network link costs increase, while per capita publication outputs decrease (Hsiehchen, Espinoza and Hsieh 2015).

The National Academies report (National Research Council 2015) recommended pre- and post-assessments to measure team effectiveness, leadership structures that foster diversity in task management, and supportive institutional policies. Hinrichs et al. (2017) added to these recommendations, highlighting the importance of real-time assessment, attention to relational management, and internal loci (individual autonomy and purpose). Other essential team ingredients include a shared mission and mental model, team trust and cohesion, attending to communicative processes, conflict resolution processes, and collaborative leadership (Börner et al. 2010; Bennett and Gadlin 2012; Baker 2015).

2.2 Approaches used to evaluate team science

Case studies, bibliometric analysis, and mixed methods evaluations have been used to provide summative insights about successes and failures of interdisciplinary initiatives (Gazni and Thelwall 2016; Tsai, Corley and Bozeman 2016; Willis et al. 2017). Social network analyses have shown changes in cross-disciplinary publication patterns (Wu and Duan 2015), and new tools are being developed to measure disciplinary diversity within project and coauthor teams (Aydinoglu, Allard and Mitchell 2016; Bergmann et al. 2017).

Less has been published about formative evaluation in team science initiatives. Since team science projects can face intense scientific and collaborative challenges during implementation, formative evaluation approaches may contribute to team learning, knowledge integration, and adaptive management, potentially strengthening research quality and outcomes (Bergmann et al. 2005; Stetler et al. 2006; Belcher et al. 2016). Participatory evaluation processes can further bolster interdisciplinary readiness and positively influence team performance (Trochim et al. 2008). Process measures can also be used to examine team functioning (Strasser et al. 2010) and identify factors that influence team performance and collaborative outcomes (Klein 2008).

In ‘mode 2’ research, stimulating co-creation of knowledge and the mobility of knowledge across the researcher-decision maker

divide involves team embedding (Contandriopoulos et al. 2010), which purposefully situates researchers in either physical or virtual proximity to knowledge users. Since this embedding approach may illuminate contextual and practical considerations pertinent to knowledge utilization, it may be also be relevant to team science evaluations. However, we were unable to find published examples of team embedding for the purposes of formative evaluation.

3. The formative evaluation case: REPOPA project

The REPOPA project¹ ran from 2011 to 2016 and was funded by the European Commission. It examined best practices for evidence-informed physical activity policymaking in six European countries: Denmark, Finland, The Netherlands, Italy, Romania, and the UK (Aro et al. 2016). Designed as a ‘mode 2’ transdisciplinary team science project, REPOPA implemented linked studies that tested innovative cross-sectoral approaches to and developed indicators for evidence-informed local-level policymaking (Aro et al. 2016). Since REPOPA intended its activities to support the use of research evidence by policymakers, European study settings with distinctive health policy-making processes and policy–research interfaces were selected. Knowledge translation² approaches emphasized a collaborative, reciprocal process of knowledge-sharing involving researchers, policymakers, and other stakeholders (Aro et al. 2016). Local policy stakeholders were involved in the project as research participants, received reports on project findings, and were engaged in project networks and national platforms.

The project consortium (hereafter referred to as the ‘consortium’) consisted of over 30 researchers representing 10 organizations (academic institutions, research organizations, and policy-related units) from the six European study countries and Canada. English was the consortium’s common language of communication but was a national language for only two countries. The REPOPA project comprised seven work packages including the evaluation work package and was deliberately structured to stimulate research synergy and crossover learning by establishing work package teams that crossed country and organizational lines.

3.1 Embedding the evaluation in the REPOPA case project

The authors of this manuscript led the evaluation work package; N. E. and S. R. were embedded in the REPOPA project for its duration and involved in all its phases. The consortium’s decision to embed the evaluators reflected its desire to mirror the collaborative, evidence-informed, contextually informed approach used in REPOPA’s research.

We negotiated the boundaries of our embedded role over the first 2 years of the project through multiple discussions with the project coordinator, work package leaders, and at annual consortium meetings. While evaluation was our principal focus, three of us (N. E., S. R., and S. V.) also had research roles on other REPOPA work packages. The evaluation lead investigator (N. E.) had some involvement as a coinvestigator in all research work packages; S. R. (evaluation colead) contributed to research discussions with three work packages and project dissemination meetings; and S. V. participated in early phases of one work package.

The consortium agreed to a clear division of responsibility between production and use of formative evaluation findings. The evaluation team was responsible for providing monitoring assessments and

related action recommendations (suggested project action steps arising from these evaluation findings), while the consortium as a whole was responsible for decisions about utilizing findings.

The formative evaluation was grounded in principles of collaborative engagement (Piggot-Irvine and Zornes 2016) and designed to be utilization-focused (Patton 2000). We identified consortium members (the research team), rather than the project funder or external stakeholders, as intended beneficiaries and owners of formative evaluation efforts. During the first year of the project, we involved the consortium in several rounds of consultation to codevelop and refine the formative evaluation strategy. These discussions brought to light three underlying concerns: (1) annual formative evaluation data collection should not become an additional and onerous burden for consortium members; (2) scientific work activities were the responsibility of work package leaders and outside formative evaluation boundaries; and (3) formative evaluation outputs should be practical and pragmatic. The consortium agreed that the purpose was to inform iterative decision-making about optimizing resources (time and strategies) to enhance the project's ability to deliver on research outcomes and impacts.

Henceforth in this article, we will use the term formative evaluation to indicate both formative and embedded evaluation as characterized our approach.

3.2 REPOPA formative evaluation: Aims and principles

The REPOPA formative evaluation used the Plan-Do-Study-Act cycle (Langley et al. 2009), designed to monitor internal project processes, research implementation and dissemination plans, and the consortium's scientific and policy networks. We aimed to provide external points of reflection about successes and challenges, maximize interdisciplinary team learning, and thereby add value to REPOPA's implementation.

We worked with the consortium to formulate guiding principles, identify potential indicators, and establish feedback mechanisms between the evaluation team and the rest of the consortium. Three central principles were agreed: the formative evaluation would use a participatory approach, regularly solicit input from consortium members, and deliberately encourage consortium utilization of findings to strengthen project decision-making about adjustments to work plans, methodological approaches, and team processes. We committed to provide annual, targeted feedback to the consortium reflecting the diversity of contexts and settings where researchers were working and to stimulate the consortium's timely consideration and use of relevant evaluation findings.

4. REPOPA longitudinal case study methods

4.1 Design

We used a descriptive, longitudinal single case study design (Yin 2009; Gustafsson 2017). Our case comprised the formative evaluation of the REPOPA project (described above); research objectives were to understand the acceptability and utility of this formative and embedded evaluation approach and describe whether and how formative results and related recommendations strengthened collaboration among team members.

The case study approach is particularly suited to generating an in-depth understanding of complex interventions in their real-life context (Crowe et al. 2011). REPOPA's context involved researchers and interventions in six national settings. Policy-researcher interfaces ranged from top-down national systems, where researchers

had nascent policymaker networks, to more interactive systems, where researcher-policymaker engagement was both expected and well established. Our case included interactions among consortium members and consortium perspectives on its engagement with stakeholders. However, assessing external stakeholders' perspective on project scientific work was outside the boundaries of the case.³

4.2 Participants

All consortium members (project coordinator, secretary, and all research team members at European sites) were invited to participate in the annual data collection process. The sample was renewed each year in the month before data collection to reflect changes among consortium members.

4.3 Ethics approval

Ethics approval for evaluation activities was received from the University of Ottawa Research Ethics Board.

4.4 Data collection

Quantitative and qualitative data were collected during four annual evaluation cycles (2013–16) and supplemented by additional data sources.⁴ Quantitative questionnaires were administered via Survey Monkey and asked participants to consider their REPOPA project experiences over the previous year. Qualitative interviews were conducted via Skype.

A team collaboration questionnaire examined satisfaction with project collaboration, and communication and engagement in knowledge translation activities. This questionnaire included 20 items in three domains: communication (six items), collaboration (seven items), and knowledge translation (seven items). Examples of knowledge translation items were *REPOPA activities in my country have created new networks among researchers and policymakers* and *My country's context was appropriately reflected in how work packages analyzed and interpreted data*.

Internal networks among consortium members and external networks with physical activity policy stakeholders were examined using a network mapping questionnaire developed by our team. For the internal network portion, participants indicated frequency of scientific communication with each consortium member on a six-point scale. The external network portion asked respondents to list physical activity stakeholders with whom they had communicated about research or policy matters (whether related to REPOPA). They indicated frequency of contact using a five-point Likert scale (1 = *once or twice a year*; 3 = *a few times a month*; 5 = *nearly every day*) and identified stakeholder's sector, primary level of their work, and their perspective on the most important benefit received from the stakeholder connection.

Team processes were examined in annual, qualitative team interviews, organized by country in Year 1 and subsequently by work package. An interview guide with open-ended questions asked about influences of consortium diversity on collaboration, communication, and team learning; experiences with policymakers; and stakeholder engagement strategies. Interviewer field notes written after conducting interviews recorded thoughts and feelings about the process.

Consortium diversity was examined using categories of functional diversity commonly referred to as pertinent in literature on team science diversity: country, team size, institutional representation, researcher-policymaker networks, and disciplinary background. We added the category of researcher-policymaker networks, since it was relevant to REPOPA.

We provided three annual monitoring reports to the consortium, circulated 2 weeks in advance of annual face-to-face project meeting. Reports synthesized interview findings and questionnaire results and highlighted positive and problematic aspects of team. SWOT (strengths, weaknesses, opportunities, threats) analysis was used to structure interview findings from Year 2 onward. We developed action recommendations and prioritized these for consortium discussion at annual project meetings. Consortium feedback about findings and action recommendations was integrated into final versions of annual reports.

4.5 Data analysis

Interview findings examined team learnings about communication and collaboration, and collaborative influences on implementation and dissemination. We developed a categorical coding framework using inductive content analysis of the first set of annual team interviews. In subsequent years, we repeated inductive content analysis of new interviews, building on our initial coding framework.

Consortium member characteristics (disciplines, language, and organizational setting) were tabulated for each country team. Organizational settings and disciplinary backgrounds were categorized as per descriptions in original research protocol or contract amendments (when organizational affiliations changed). When individuals had graduate degrees in more than one discipline, all their disciplines were listed.

To assess acceptability of the formative evaluation, we calculated annual response rates for questionnaires (% of invited consortium members participating) and interviews (% of invited country/work package teams with at least one member participating). We also examined consortium responses to interview probes about the formative evaluation process and reviewed field notes recorded after team interviews.

To visually illustrate internal consortium communication patterns within and across the six country teams, we created sociograms using NetDraw (Borgatti, Everett and Freeman 2002). Changes in strength of internal consortium communication were compared across 4 years using average number of connections among consortium members documented in network questionnaires. External stakeholders were categorized according to sector (government, academia, and NGO) and level (local/provincial or national/international); composition was compared year to year using descriptive statistics.

To characterize action recommendations on team processes, we retrospectively reviewed all recommendations presented in annual monitoring reports. We selected those explicitly addressing team processes or team/setting diversity and inductively categorized this subset for each reporting year.

Collaboration was assessed using collaboration and network questionnaires and annual interviews. Reliability (Cronbach's alpha) of the three collaboration questionnaire subscales was assessed using data from all four data collection periods. To determine whether there had been improvements in collaboration, standardized, annual, mean scores were calculated for each subscale. Case-wise deletion was used for missing responses. One-way analysis of variance (ANOVA) was used to compare changes in scores across the 4 years of data collection.

5. Results

5.1 Response rates

Response rates for the collaboration questionnaire, network questionnaire, and interviews are presented in Figure 1. Rates were consistently high, ranging from 70% to 100% across the 4 years of data collection.

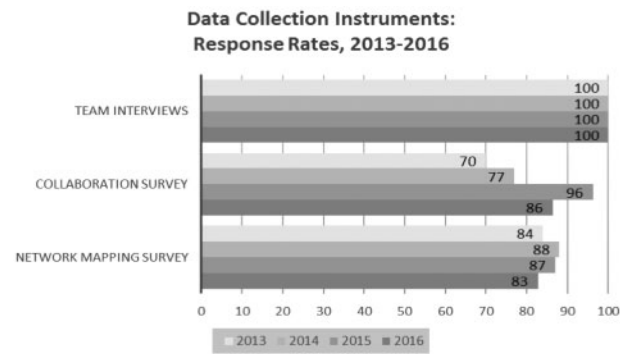


Figure 1. Annual response rates for data collection instruments (2013–16).

A total of 30 interviews were conducted over 4 years, majority ($n=23$) as work package or country team interviews with several ($n=7$) individual interviews. Interviews averaged 59 min. After the first year, some members participated in more than one interview due to involvement in more than one work package.

5.2 Characteristics of consortium members

All work package teams included a mix of researchers from different scientific traditions and linguistic, organizational, and policy contexts. Work package teams ranged in size from 3 to 24 people, while country teams ranged from 1 to 11 people (see Table 1). Some country teams were involved in as few as three work packages, others in as many as seven. All but one country team included researchers with a public health background. There were 14 scientific disciplines represented on the consortium. The nine European institutions in the project were a mix of academic institutions ($n=3$), nonuniversity research institutes ($n=4$), and practice-oriented organizations ($n=2$); each had different mandates for bridging researchers and policymakers.

5.3 Consortium collaboration and communication

Standardized Cronbach's alphas for the three team collaboration subscales (communication, collaboration, and knowledge translation) were 0.81, 0.88 and 0.91, respectively. Small, year-to-year differences in mean scores were seen for each subscale. An improvement in scores was observed for both communications and collaboration subscales between the first and second years of the project. Measures of collaboration and knowledge translation were lowest in Year 3; both scores improved in Year 4. ANOVA results are presented in Table 2. There were no statistically significant differences in subscale scores across the 4 years of reporting.

Communication patterns were also determined from responses to internal network questionnaire (see network maps in Figure 2). In Year 1, communication was primarily country-based; more frequent contacts occurred among members in same country team and also between each country team and the management team. In Years 2 and 3, coinciding with rollout of multiple research work packages, communication among country teams increased. However, these latter connections appeared weaker in the final year. In the first year, consortium members averaged 19 connections each, or 70.6% of all possible connections. By Year 3, average number of connections had increased to 26. However, average number of connections was lowest (18.7) in Year 4.

Table 1. Characteristics of REPOPA Consortium members by country

Country ^a	Size of country team (range in number of members during project)	Number (type) of institutions represented on team during project	Focus of institution(s)	Number of work packages that team was involved in ^b	Team experience with researcher–policy-maker networks	Team members' disciplinary backgrounds ^c
A	Large(10–11)	2 (1 university and 1 research center)	Health promotion; prevention and health	7	Extensive especially at municipal level	Psychology Public health Sports science Medicine Epidemiology Nutrition science
B	Small(2)	1 (government research institute)	Research for decision-making on health	3	Extensive especially with national ministry	Public health Sociology Health education
C	Medium (4–6)	3 (1 university and 2 practice-oriented organizations)	Bridging health science and practice; health promotion practice	5	Extensive through prior projects	Public health Health promotion Nutrition science
D	Medium (4–5)	1 (university)	Public health research	4	Limited	Medicine Public health Political science Communication
E	Medium (4–5)	1 research institute	Science communication	4	Some but not in physical activity	Law Communication Physics Cultural Anthropology
F	Small (1–3)	0–1* (primary care trust)	Public health practice	3	Extensive but not in physical activity	Public health Dentistry

^a Excludes Canada.

^b The original institutional partner was unable to continue with the project after year 2. Specific research activities were later sub-contracted to a consultant who had been involved earlier in the project.

^c Some consortium members represented more than one disciplinary background.

5.4 Networks with external stakeholders

Table 3 shows networks between consortium members and external policy stakeholders. From a high of 83 total external connections in the first year, the number dropped in each subsequent year to a low of 30 in the final year. Stakeholder sector remained consistent across project years; they were predominantly from government (range: 78.4–73.3%), followed by academia (range: 12–16.7%), and non-governmental organizations (range: 9.6–10%). However, over the course of the project, there was a marked change in the system level at which stakeholders worked. In 2013, 64.1% of identified stakeholders were local/provincial and 35.9% national/international. By 2016, this composition had reversed: 62.8% were national/international stakeholders and 37.2% local/provincial.

Consortium members reported a shift in primary benefits they received from their external stakeholders during the project (see Table 3). The largest differences between Years 1 and 4 were seen for three types of benefits listed on the questionnaires: *help me understand the policy context* (52.9% versus 20.8%), *help me disseminate my research findings* (4.8% versus 29.2%), and *give feedback on scientific aspects of my work* (2.8% versus 12.5%).

5.5 Consortium perspective on collaboration and diversity

Interviewees indicated several ways that diversity of scientific disciplines and languages influenced research discussions. Participants described having to address these differences to design their work package interventions, finalize methodological approaches, and develop joint understanding of concepts central to research on evidence-informed policy-

making. One participant described this as ‘a very delicate phase of the REPOPA project’ which was challenging ‘because we actually had to find agreement on many, many points [in order to] synthesize results and inputs from many work packages’ (2015 team interview).

Participants indicated that investing time in this integrative process ultimately benefited research goals, though sometimes creating implementation delays.

One of the biggest challenges [was] to try to synchronize things... adapting the [intervention] to the different context, and then in the phase of data analysis with the questionnaires and the observation data... This is inherent to the work that we're doing. For the future... be conscious... [of] how much time this takes and how much effort. (2016 team interview)

Consortium members spoke about challenges arising from varied backgrounds and perspectives within work packages and the project as a whole:

So it is really a challenge to coordinate different disciplines, different ambitions, different collaborative willingness... so that we could find the common direction, level of ambition in the work. (2016 team interview)

However, consortium members appreciated the potential for cross-learning among researchers with different bodies of expertise and from different institutions.

[It's] a great learning opportunity to have several institutions involved in [work package X]... we've been able to learn from each other and supplement... each other's knowledge... On the other hand, ... it has

Table 2. Collaboration survey: summary of ANOVA for comparison of mean scores per year for each subscale

Subscale	Mean score (SD)				Sum of squares	Mean square	F-value	P-value
	2013 N = 21	2014 N = 20	2015 N = 26	2016 N = 19				
Communication	4.68 (0.84)	4.89 (0.80)	4.80 (0.60)	4.87 (0.71)	20.45	6.82	0.35	0.792
Collaboration	4.64 (0.98)	4.86 (0.71)	4.57 (0.81)	4.80 (0.81)	62.94	20.98	0.61	0.612
Knowledge translation	4.75 (0.72)	4.69 (1.02)	4.46 (0.82)	4.85 (0.90)	95.16	31.72	0.87	0.459

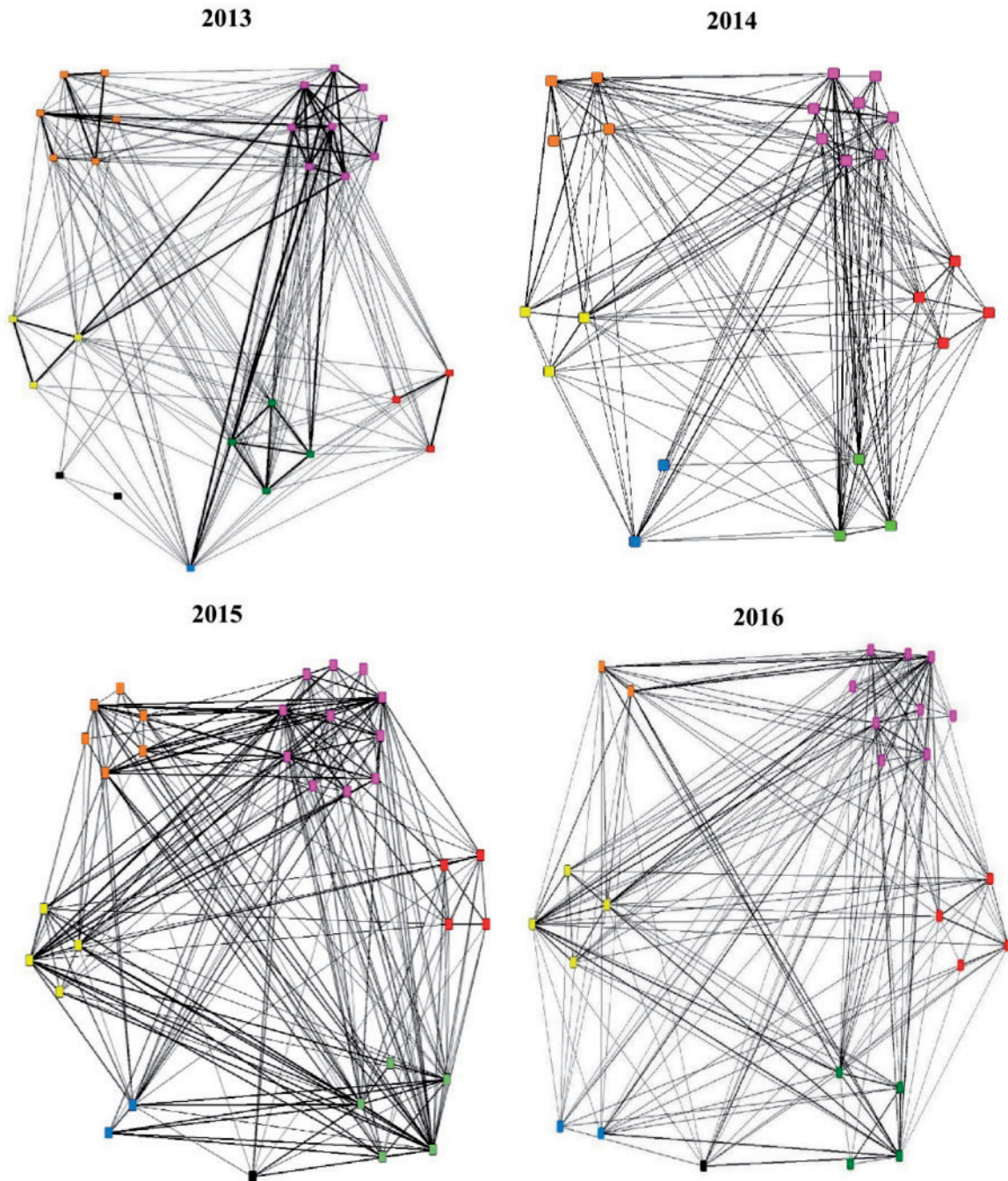


Figure 2. Internal consortium networks, 2013–16, showing connectedness of country teams.

Table 3. Networks between REPOPA members and external policy stakeholders, 2013–16

Profile category	Year			
	2013 (n = 29) N (%)	2014 (n = 24) N (%)	2015 (n = 29) N (%)	2016 (n = 22) N (%)
Average number of contacts per Consortium member	3.96	4.24	5.64	4.80
Primary sector of impact				
Government	65 (78.4)	49 (72.1)	49 (75.4)	22 (73.3)
Academia	10 (12.0)	10 (14.7)	11 (16.9)	5 (16.7)
Non-governmental	8 (9.6)	9 (13.2)	5 (7.7)	3 (10.0)
Total	83	68	65	30
Level of impact				
Local (city) or province/state	50 (64.1)	47 (69.1)	32 (49.2)	16 (37.2)
National/international	28 (35.9)	21 (30.9)	33 (50.8)	27 (62.8)
Total	78	68	65	43
Primary benefit to REPOPA member				
Helps me understand the policy context for my work	55 (52.9)	24 (33.3)	15 (19.0)	10 (20.8)
Helps me disseminate my research findings	5 (4.8)	14 (19.4)	16 (20.3)	14 (29.2)
Helps me identify other stakeholders	11 (10.6)	2 (2.8)	0 (0.0)	2 (4.2)
Provides access to decision makers	10 (9.6)	5 (6.9)	2 (2.5)	6 (12.5)
Helps me gain political support for my research findings	4 (3.8)	1 (1.4)	6 (7.6)	3 (6.3)
Gives me feedback on scientific aspects of my work	4 (3.8)	6 (8.3)	15 (19.0)	6 (12.5)
Other	11 (10.6)	16 (22.2)	19 (24.1)	2 (4.2)
Multiple options selected	2 (1.9)	0 (0.0)	7 (8.9)	5 (10.4)
Missing response	2 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)
Total	104	68	80	48
Total number of stakeholders	107	82	86	48

been a challenge to be engaged... [in this for] myself and our institution because... we didn't have experience... within that research area that [work package X] covers. (2014 team interview)

Participants sometimes used the interview context to air interpersonal or methodological conflicts:

...In the spirit of transparency I felt that having this [conflict] out loud, it's useful to the process. And also it's lessons learned for ... people that will implement projects to know that things aren't always smooth... there are things that need to be settled, need to be discussed, adjusted. [2015 team interview]

Our field notes indicate when team conflicts were raised by interviewees, we carefully maintained our evaluator role and were not drawn into conflict resolution, even when the latter was requested.

5.6 Stakeholder networks

Institutions where researchers were employed and country teams had different preexisting networks they could draw on. Consortium members who lacked preexisting networks with policy stakeholders identified this as a challenge:

This [physical activity focus] has been a great challenge for us... because we went much more from a social status of science perspective than [a] public health perspective... It has been really very challenging to be connected with existing groups and build networks in this field. (2016 team interview)

In other settings, teams could more easily tap into and grow preexisting networks:

Our network of researchers, policy makers, both on local level and on national level has been growing from the beginning and

now, especially with work package X [activities]... it flows all together. (2016 team interview)

REPOPA's interventions involved policy actors in six countries (Aro et al. 2016; Bertram et al. 2016; Spitters et al. 2017). These country systems varied in how policy stakeholder authority, autonomy, and responsibility were configured and operationalized. Teams had to comprehend divergent setting-specific expectations to align research and knowledge translation efforts with locally relevant stakeholder approaches to collaboration.

As one participant noted:

Even though the settings might seem similar, the stakeholders and the needs are different so the [intervention] has developed differently... (2014 team interview).

Participants also described uncertainties inherent in conducting research in real-life contexts and with partners outside of academia, indicating that this added to implementation complexity. Health system governance and economic changes in some partner countries during the project had an impact especially on those from practice- and policy-oriented organizations compared to those from universities. In one country, the practice-based organization was unable to remain in the project after health system restructuring. In another country, the university-based team twice had to rebound from successive bankruptcies of their community-based research partner; the fallout involved higher workloads and fewer resources for remaining team members and the loss of some direct stakeholder connections.

We have more research and implementation projects where we work together with [those] in the healthcare field and in welfare field... It's very good to work in the real life setting... but it always also generates more difficulties in the research

Table 4. Selected examples of action recommendations in monitoring reports, by focal area (2013–15*)

Emergent focal area for recommendation	2013 (first monitoring report)	2014 (second monitoring report)	2015 (third monitoring report)
Team processes for communication and collaboration	Define team roles and responsibilities for different types of members (researchers, trainees, and project staff) and organizations	Use upcoming face-to-face annual meeting to launch work package X and discuss implementation plans	Build momentum for final project symposium through periodic consortium-wide Skype meetings
Develop strategic responses to strengthen cross-project approaches	Identify and implement strategies to support potentially vulnerable country or work package teams: e.g. expand small teams, proactive communication	Draw on richness of country/organizational contexts, scientific expertise, and dissemination experiences to support cross-consortium learning and problem-solving	Develop post-project vision and plan for publications and future collaborations
Use consortium diversity to enhance country strategies	Consolidate lessons learned within and across countries about effective strategies for recruiting and engaging policy stakeholders	Develop country-specific targeted knowledge translation plans. Hold cross-country Skype sessions to share strategies and joint problem-solving	Articulate country goals and work plans to achieve key knowledge products in final year
Use consortium diversity to inform novel scientific learning	Actively share local experimentation by country teams; proactively generate cross-jurisdictional learning to expand REPOPA's reach and add value to cross-country collaboration	Consider writing a manuscript that compares and contrasts lessons learned from diverse settings about tailoring interventions to context, recruiting stakeholders, and developing knowledge products	Identify propositions that could be examined through an integrative analytic approach involving more than one project and more than one country

Notes: While data were collected across four time periods, action recommendations were included in the three annual monitoring reports provided to the consortium. No action recommendations were included in the project's final summative evaluation report.

projects...It's very difficult to plan ahead for three, four, five years and [assume] that things will stay the same. (2016 team interview)

5.7 Action recommendations to the consortium

Four different focal areas for action recommendations were identified over the course of the project: team processes for communication and collaboration, developing strategic responses to strengthen cross-project approaches, using consortium diversity to enhance country strategies, and using consortium diversity to inform novel scientific learning. Selected examples of these focal areas are presented in Table 4. While annual action recommendations covered all four focal areas, specific recommendations changed year to year in response to evolving consortium dynamics, work package progress, and changes in project opportunities and threats.

In the first annual report to the consortium, many of our action recommendations aimed to help the consortium apply lessons learned about early team processes. As REPOPA progressed, opportunities emerged for work packages to capitalize on contextual differences among country settings and organizations. The final set of action recommendations emphasized targeted approaches to dissemination.

5.8 Consortium perspectives on evaluation acceptance and utility

The consortium made use of the evaluation process, findings, and action recommendations in several ways. Early on, consortium network maps were presented alongside a recommendation that work package leaders more strategically consider supports for isolated members; this immediately engaged the consortium to consider action strategies. While we had originally planned to use the network tool only in Years 1 and 4, consortium members valued being able to visualize internal connections and requested the questionnaire be administered annually. Team discussions the following year referred

to this as impetus for the more active communication pathways that had emerged within work packages.

Interview data and evaluators' field notes pointed to how interviews were used by participants as a 'safe space' to debrief and critically reflect on team dynamics, member contributions, project successes, and struggles. When new ideas about research challenges or knowledge translation approaches came up during interviews, teams sometimes followed up post-interview with their work packages to continue discussions.

Project members generally expressed satisfaction with the evaluation team's work and approach:

I have to say that I've been working for a number of years in collaborative projects across countries, and I feel that the process you've set in place for the evaluation is the best I've seen so far. At least for me, this came to be as the golden standard so far, so I think you're doing great work... it forced people to contemplate the way they were reporting to each other and the way they were cooperating with each other. (2014 team interview)

They described it as differing from previous evaluation experiences:

I told you earlier of the evaluation of [a] project published without informing the partners, but I don't think you do that. So I trust you hundred percent. (2015 team interview)

While we were considered part of the overall consortium and participated in other work packages, some participants wished that we had stepped out of our evaluation role more often:

... Sometimes I feel that you are not part of REPOPA... somehow I would like to feel like much more [of a] team, together with the [evaluation] team. (2015 team interview)

6. Discussion

This is one of the first studies examining the utility of a formative and embedded evaluation to inform and strengthen ‘mode 2’ team science. Although this is a single case study, it involved multi-country projects and researchers drawn not only from different disciplines but also from varied organizational and geopolitical settings. Results indicate that our approach, which involved applying the tenants of collaborative engagement (Piggot-Irvine and Zornes 2016), led to a set of explicit principles that guided the work, kept consortium members actively involved (as both respondents and users of data and related action recommendations) through the 4-year evaluation period, and provided new insights on managing and harnessing diversity.

We sought to mirror the project’s collaborative research approach with collaborative evaluation practice. Providing annual reports of findings alongside action recommendations enabled the consortium to respond to these ‘real-time assessments’ (Hinrichs et al. 2017). We found that evaluation processes supported consortium’s critical reflexivity and responsiveness to team functioning issues and surfaced opportunities and challenges presented by consortium diversity.

6.1 Changes in collaboration and communication

Different team issues arose as the project progressed. Interview data indicated that developing and adapting collaborative strategies was an ongoing necessity for the consortium. Early on, both collaboration questionnaire findings and interviews identified communication processes as a concern for some. Teams struggled internally in the face of differing perspectives and disciplinary miscommunication. As is common in team science endeavors, they needed to develop a common lexicon that integrated ideas from across disciplines (Fiore 2008; Read et al. 2016; Klink et al. 2017). These challenges continued as teams determined how best to tailor interventions to organizational settings and analyze findings in ways relevant to country stakeholders. All teams indicated that these processes required substantially more time than expected and caused some unanticipated delays. A tension emerged between devoting scarce consortium resources (time) into integrating diverse perspectives or pressing forward with implementation.

Perhaps because of these pragmatic tensions, quantitative collaboration findings showed no statistically significant improvements in domains of communication, collaboration, or knowledge translation over the 4 years of study. This may reflect a tendency to overestimate team coherence, communication, and collaboration ability at project outset, and increased intensity of work, necessitated by contractual funding deadlines, in later project phases. As consortium members reported during interviews, the real challenges of interdisciplinary and multisite work surfaced as the project progressed.

However, we did not see a decline in collaboration scores and suggest the primary reason is that various facets of team science were addressed by the consortium from the project’s beginning. As outlined in Bennett and Gadlin (2012), critical enabling factors include a shared vision, purposefully building the consortium and its diversity to enhance collaboration readiness (Hall et al. 2008), managing conflict, and setting clear expectations for sharing credit and authorship. The consortium tackled these factors in a number of ways, including developing consortium publication guidelines, extending members’ understanding of the differing study contexts by rotating annual project meetings among countries, and

examining diversity as it pertained to ethics approvals (Edwards et al. 2013). The embedded evaluation process helped surface and reinforce these strategies to strengthen team science.

6.2 Changes in consortium networks

Networks warrant particular investigation in team science (Rivera, Soderstrom and Uzzi 2010; Okamoto 2015; Davids & Frenken 2018). We identified changes in the project’s internal network that seemed to track closely with implementation progress and proximity. Initially, researchers communicated most frequently with others from the same country team. This is consistent with findings by Norman et al. (2011) and Hardeman et al. (2015) showing that physical colocation influences communication patterns. As work package activities cutting across countries intensified, communication ties among countries strengthened. However, it was in the third rather than final year when communication networks were strongest, likely because researchers were then most intensively engaged in multiple simultaneous work packages. In the final year, researchers’ priorities shifted to translating and disseminating project findings to local contexts. These shifts in the work foci for the consortium over the course of the project are particularly apparent in their changing descriptions of year-to-year primary benefits received from external stakeholders.

In early years, project research took place independently in study countries (e.g. policy analysis), while joint activities (e.g. policy simulation workshops) were more frequent in later years. Work package members also used common attendance at international conferences to hold team meetings. So over time, there were fluctuations in physical proximity of team members vis-a-vis the work. As teams successfully worked together over the years, cognitive and social proximity likely increased. This ebb and flow in internal and external project networks that occurred in relation to project priorities demonstrates consortium efforts to strengthen internal networks and vitalize external networks to achieve project goals.

6.3 Implications of diversity

Internal consortium diversity offered the project both strengths and challenges; the formative evaluation helped the consortium respond as these arose. We identified two kinds of diversity: deliberate diversity (aspects which the consortium had purposefully designed into the structure of work package teams from project outset) and emergent diversity (arising from synergistic interactions among settings and organizations).

While there is a considerable literature on understanding how multiplicity of disciplinary backgrounds on a research team is integral to collaborative research (Bennett & Gadlin 2012; National Research Council 2015; Van Noorden 2015), less has been published about the influence of other aspects of diversity. In our evaluation, the interactions of three types of diversity stood out: those of disciplinary backgrounds, institutional settings, and geopolitical settings. Each work package team represented a mix of these diversities, which increased work effort and intensity, since consortium members had to bridge differing perspectives about and levels of familiarity with varying methodological approaches, systems for policy–researcher interface, and institutional priorities. Diversity affected approaches to knowledge generation and translation when researchers had to address differing expectations within their own workplaces (such as preferences for peer-reviewed publications versus other types of oral or written dissemination). Differences in

whether networks between academics and policymakers were well established or nascent, and the types of links that were considered legitimate and supported by their institutions influenced research and dissemination processes and strategies. A third realization was that diversity of experiences, settings, and disciplinary perspectives was a project asset, providing, for example, access to wider and more diverse stakeholders within and among countries.

6.4 Engagement in evaluation processes

Similar to Patton (2008), consortium members reported that the formative evaluation approach contributed to their engagement with evaluation findings and action recommendations. Consistently high response rates to evaluation tools across all 4 years of data collection and provision of adequate time for strategic discussions of evaluation findings at annual meetings suggest the consortium considered these investments of their time to be worthwhile. Consortium requests for more frequent network data collection were also indicative of consortium engagement in the evaluation process. However, we were sometimes constrained in how fully we could support the consortium's intention to follow up on particular findings, for example when ethical constraints prevented identification of isolated individuals in the consortium network.

While much of the consortium was not initially familiar with participatory and embedded evaluation, we were able to demonstrate what this looked like in practice. Essential to the conduct of the formative evaluation was active consortium involvement not only in evaluation design but also in iterative, critical reflection on action recommendations and a cross-project problem-solving orientation to addressing contextual and implementation challenges, indicating they viewed the evaluation process as practical, timely, and relevant for project decision-making.

Focusing consortium attention on internal communication and team processes by sharing findings and recommendations supported the consortium in identifying and strengthening collaborative mechanisms to help manage diversity. Other published descriptions of making information available to project teams suggest collaborative research strategies that were similar to ours. For example, Klink et al. (2017) included a process assessment of team functioning for an interdisciplinary climate risk project; Prokopy et al. (2015) used a team survey in the same project to improve team communication about research products; and DeLorme et al. (2016) described evaluation processes to support collaborative engagement mechanisms and project planning in their study of coastal sea-level rise.

Reflexivity and feedback processes have been shown to positively influence team decision-making and outcomes (Eddy, Tannenbaum and Mathieu 2013). Our overall evaluation process, and particularly the interviews, seemed to support the practice of reflexivity (Vindrola-Padros et al. 2017) among team members. It is our impression that participants commented openly about their own and other work package teams, in part because interviews were conducted by known and trusted consortium members who were not involved with country implementation.

While it is premature for us to attribute stronger team science to the embedded evaluation, there are early indications that it contributed to the project. We found that working with the consortium as embedded evaluators and as full project members built consortium trust, reflexivity, and engagement in formative evaluation processes, which in turn enhanced the relevance and utilization of our findings

and recommendations. High response rates, positive assertions about the evaluation process, and responsiveness to action recommendations were indicative that evaluation feedback was a valued input to team functioning. However, our quantitative collaboration measures were not responsive to these changes.

6.5 Embedded evaluation approach and team science

Our approach was consistent with characteristics of embedded research identified by Vindrola-Padros et al. (2017). We maintained dual affiliation to research work packages on the one hand, and the evaluation work package on the other hand, and were considered part of the consortium. This embedded participation was a topic of ongoing reflection by evaluation team and the rest of the consortium, as we needed to determine whether we would be embedded only as evaluators or in both evaluation and coinvestigator roles. To avoid conflict of interest, the consortium agreed at the outset that scientific evaluation of work package activities would take place during the summative, not formative, evaluation.

Because our evaluation team was based outside of Europe, there was also a tendency to engage us differently in scientific activities; we provided a reference point and conduit to pertinent research and experiences in Canada. In contrast, European colleagues were necessarily involved in all facets of the research including meeting the many operational and management challenges essential to rolling out a large research project and engaging directly with local stakeholders. We revisited our roles annually with the consortium. When team tensions were identified during interviews, we reminded the project coordinator of our evaluation role. Our collaboration as coinvestigators varied with the progression of research work packages, intensifying during the project's final year with end-of-project dissemination plans and activities.

In our experience, embedding our evaluation team throughout project life cycle meant we gained deeper familiarity with work packages, study contexts, and implementation challenges. Being embedded facilitated our ability to surface potential learnings arising from diversity and to adapt interviews to emerging areas of consortium concern. While some researchers have mentioned a lack of scientific objectivity as a potential limitation for embedded evaluators (Marshall 2014), we think this was offset by the opportunity to proactively collaborate in research knowledge generation and translation activities and experience first hand the project tensions and issues that arose and to see how they were resolved.

Our evaluation approach blended the strengths of participatory and utilization-focused evaluation. The former can bolster interdisciplinary readiness and positively influence team performance (Stokols et al. 2008), while the latter may facilitate implementation responsiveness (Flowers 2010). We agree with Vindrola-Padros et al. (2017) that codeveloping evaluation guidelines in the early stages of the project provided an important touchstone as consortium work unfolded. Other contributing factors included actively engaging the consortium in evaluation design, creating feedback processes on the monitoring report, and respecting anonymity and confidentiality requirements.

7. Limitations

While our use of a single case study allowed for in-depth understanding of interactions between evaluation processes and team

science within the REPOPA project, it limits generalizability (Gustafsson 2017).

Our measure of team collaboration may not have been optimal. Although its reliability was good, scores at time one were high, leaving limited room for improvements. This is a conundrum facing those measuring team collaboration, since team members may tend to overestimate team coherence, communication, and collaboration abilities at project outset.

Only the perspectives of consortium members were sought. Had we included external stakeholders as participants in the evaluation, this may have informed improvements in collaboration, particularly in relation to knowledge translation and dissemination strategies.

8. Future research

We suggest several areas for future research. While our formative and embedded evaluation approach seems promising in terms of acceptability, utility, and team engagement, determining whether embedded evaluation strengthens team science will require a longitudinal experimental or quasi-experimental study. We suggest comparing teams that are purposefully diverse in three parameters (discipline, and organizational and geopolitical settings) and which do or do not have embedded evaluators. Complementary interventions could focus on strengthening team science, using the growing body of preparedness resources and field guides (Bennett, Gadlin and Levine-Finley 2010; Vogel et al. 2013).

Perspectives of external stakeholders who interface with interdisciplinary research team members are needed to broaden understanding of what constitutes team science at the knowledge production–knowledge use interface.

Finally, robust measures are needed that are more sensitive to change over time in the phases of collaborative processes and include items assessing trust among team members, shared vocabulary lexicon, the role of boundary spanners, and interdisciplinary conflict management.

9. Conclusion

This embedded, mixed methods, formative evaluation approach to examine team science efforts is promising with respect to its utility, acceptability, and researcher engagement. Although we did not find significant improvements in team collaboration over the four periods of data collection, collaboration levels remained high even during the challenging ‘slogging through’ phases of the project. Three types of diversity stood out in this multi-country project—disciplinary, organizational, and geopolitical. All need to be taken into account, including interactions among them, in future evaluations.

An embedded evaluation strategy, where evaluators are part of the project team, is a novel approach to formative team science evaluations. Success depends on a trusting, collaborative relationship between evaluation team members and other project members. This can be enhanced by: engaging the entire project in identifying team science objectives for the evaluation at the outset, codeveloping guiding principles, and encouraging team reflexivity throughout the evaluation.

Notes

1. While REPOPA’s scope (involving a set of conceptually linked interdependent studies) may classify it as a program of research, we follow the European Union terminology in referring to it as a project.

2. Knowledge translation was defined by REPOPA as: “knowledge integration,” two-way knowledge sharing so that researchers and policymakers as research users are engaged in the entire research process (Jansen et al. 2012). In the REPOPA interventions this means that researchers and policymakers jointly in their endeavor develop policies which are salient and relevant to policymakers and to the problems’ (Aro et al. 2016)
3. We were also responsible for designing and conducting the summative evaluation, which included assessing the scientific impact of the REPOPA Project; this will be reported on in a separate publication.
4. The authors may be contacted for more details about data collection and analysis methods.

Acknowledgement

The authors thank all members of the REPOPA consortium for their participation in evaluation activities and discussions.

Members of the REPOPA Consortium (current composition): Coordinator: University of Southern Denmark (SDU), Denmark: Arja R. Aro, Maja Bertram, Christina Radl-Karimi, Natasa Loncarevic, Gabriel Gulis, Thomas Skovgaard, Ahmed M Syed, Leena Eklund Karlsson, and Mette W Jakobsen. Partners: Tilburg University (TiU), The Netherlands: Ien AM van de Goor, Hilde Spitters; the Finnish National Institute for Health and Welfare (THL), Finland: Timo Ståhl and Riitta-Maija Hämäläinen; Babes-Bolyai University (UBB), Romania: Razvan M Chereches, Diana Rus, Petru Sandu, and Elena Bozdog; the Italian National Research Council (CNR), Italy: the Institute of Research on Population and Social Policies (IRPPS): Adriana Valente, Tommaso Castellani, and Valentina Tudisca; the Institute of Clinical Physiology (IFC): Fabrizio Bianchi and Liliana Cori; School of Nursing, University of Ottawa (uOttawa), Canada: Nancy Edwards, Susan Roelofs, and Sarah Viehbeck; and Research Centre for Prevention and Health (RCPH), Denmark: Torben Jørgensen, Charlotte Glümer, and Cathrine Juel Lau.

Funding

This work, within the REsearch into Policy in Physical Activity (REPOPA) Project, October 2011–September 2016, was supported by the European Union Seventh Framework Programme (FP7/2007–13); grant agreement number 281532. This document reflects only the authors’ views, and neither the European Commission nor any person on its behalf is liable for any use that may be made of the information contained herein.

References

- Aro, A. R. et al. (2016) ‘Integrating Research Evidence and Physical Activity Policy Making—REPOPA Project’, *Health Promotion International*, 31/2: 430–9.
- Aydinoglu, A. U., Allard, S., and Mitchell, C. (2016) ‘Measuring Diversity in Disciplinary Collaboration in Research Teams: An Ecological Perspective’, *Research Evaluation*, 25/1: 18–36.
- Baker, B. (2015) ‘The Science of Team Science: An Emerging Field Delves into the Complexities of Effective Collaboration’, *BioScience*, 65/7: 639–44.
- Belcher, B. M. et al. (2016) ‘Defining and Assessing Research Quality in a Transdisciplinary Context’, *Research Evaluation*, 25/1: 1–17.
- Bennett, L. M., and Gadlin, H. (2012) ‘Collaboration and Team Science: From Theory to Practice’, *Journal of Investigative Medicine*, 60/5: 768–75.
- Bennett, L. M., Gadlin, H., and Levine-Finley, S. (2010) *Collaboration and Team Science: A Field Guide*. Bethesda, MD: NIH National Cancer Institute.
- Bergmann, M. et al. (2005) *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*. Frankfurt, Germany: Institute for Social-Ecological Research (ISOE) GmbH.

- Bergmann, T. et al. (2017) 'The Interdisciplinarity of Collaborations in Cognitive Science', *Cognitive Science*, 41/5: 1412–18.
- Bertram, M. et al. (2016) 'Planning Locally Tailored Interventions on Evidence Informed Policy Making: Needs Assessment, Design and Methods', *Health Systems and Policy Research*, 3/2: 15.
- Borgatti, S. P., Everett, M., and Freeman, L. C. (2002) *UCINET for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Börner, K. et al. (2010) 'A Multi-Level Systems Perspective for the Science of Team Science', *Science Translational Medicine*, 2/49: 49cm24.
- Boschma, R. (2005) 'Proximity and Innovation: A Critical Assessment', *Regional Studies*, 39/1: 61–74.
- Cheruvilil, K. S. et al. (2014) 'Creating and Maintaining High-Performing Collaborative Research Teams: The Importance of Diversity and Interpersonal Skills', *Frontiers in Ecology and the Environment*, 12/1: 31–8.
- Cheung, S. Y. et al. (2016) 'When and How Does Functional Diversity Influence Team Innovation? The Mediating Role of Knowledge Sharing and the Moderation Role of Affect-Based Trust in a Team', *Human Relations*, 69/7: 1507–31.
- Contandriopoulos, D. et al. (2010) 'Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature', *The Milbank Quarterly*, 88/4: 444–83.
- Cousins, J. B., Whitmore, E., and Shulha, L. (2013) 'Arguments for a Common Set of Principles for Collaborative Inquiry in Evaluation', *American Journal of Evaluation*, 34/1: 7–22.
- Crowe, S. et al. (2011) 'The Case Study Approach', *BMC Medical Research Methodology*, 11: 100.
- Cummings, J. N., and Kiesler, S. (2007) 'Coordination Costs and Project Outcomes in Multi-University Collaborations', *Research Policy*, 36/10: 1620–34.
- Dauids, M., and Frenken, K. (2018) 'Proximity, Knowledge Base and the Innovation Process: Towards an Integrated Framework', *Regional Studies*, 52/1: 23–34.
- de Jong, S. et al. (2011) 'Evaluation of Research in Context: An Approach and Two Cases', *Research Evaluation*, 20/1: 61–72.
- DeLorme, D. E. et al. (2016) 'Developing and Managing Transdisciplinary and Transformative Research on the Coastal Dynamics of Sea Level Rise: Experiences and Lessons Learned', *Earth's Future*, 4/5: 194–209.
- Eddy, E. R., Tannenbaum, S. I., and Mathieu, J. E. (2013) 'Helping Teams to Help Themselves: Comparing Two Team-Led Debriefing Methods', *Personnel Psychology*, 66/4: 975–1008.
- Edwards, N. et al. (2013) 'Challenges of Ethical Clearance in International Health Policy and Social Sciences Research: Experiences and Recommendations from a Multi-Country Research Programme', *Public Health Reviews*, 34/1: 11.
- Falk-Krzesinski, H. J. et al. (2010) 'Advancing the Science of Team Science', *Clinical and Translational Science*, 3/5: 263–6.
- Fiore, S. M. (2008) 'Interdisciplinarity as Teamwork: How the Science of Teams Can Inform Team Science', *Small Group Research*, 39/3: 251–77.
- Flowers, A. B. (2010) 'Blazing an Evaluation Pathway: Lessons Learned from Applying Utilization-Focused Evaluation to a Conservation Education Program', *Evaluation and Program Planning*, 33/2: 165–71.
- Gazni, A., and Thelwall, M. (2016) 'The Citation Impact of Collaboration between Top Institutions: A Temporal Analysis', *Research Evaluation*, 25/2: 219–29.
- Gibbons, M. (2000) 'Mode 2 Society and the Emergence of Context-Sensitive Science', *Science and Public Policy*, 27/3: 159–63.
- Goring, S. J. et al. (2014) 'Improving the Culture of Interdisciplinary Collaboration in Ecology by Expanding Measures of Success', *Frontiers in Ecology and the Environment*, 12/1: 39–47.
- Gustafsson, J. (2017) *Single Case Studies Vs. Multiple Case Studies: A Comparative Study*. Academy of Business, Engineering and Science, Halmstad University, Halmstad, Sweden. <<http://www.diva-portal.org/smash/get/diva2:1064378/FULLTEXT01.pdf>>
- Hall, K. L. et al. (2008) 'The Collaboration Readiness of Transdisciplinary Research Teams and Centers', *American Journal of Preventive Medicine*, 35/2: S161–72.
- Hardeman, S. et al. (2015) 'Characterizing and Comparing Innovation Systems by Different "Modes" of Knowledge Production: A Proximity Approach', *Science and Public Policy*, 42/4: 530–48.
- Heringa, P. W. et al. (2014) 'How Do Dimensions of Proximity Relate to the Outcomes of Collaboration? A Survey of Knowledge-Intensive Networks in the Dutch Water Sector', *Economics of Innovation and New Technology*, 23/7: 689–716.
- Hinrichs, M. M. et al. (2017) 'Innovation in the Knowledge Age: Implications for Collaborative Science', *Environment Systems and Decisions*, 37/2: 144–55.
- Hsiehchen, D., Espinoza, M., and Hsieh, A. (2015) 'Multinational Teams and Diseconomies of Scale in Collaborative Research', *Science Advances*, 1/8: e1500211.
- Hüttermann, H., and Boerner, S. (2011) 'Fostering Innovation in Functionally Diverse Teams: The Two Faces of Transformational Leadership', *European Journal of Work and Organizational Psychology*, 20/6: 833–54.
- Jansen, M. W., De Leeuw, E., Hoeijmakers, M., and De Vries, N. K. (2012) 'Working at the nexus between public health policy, practice and research. Dynamics of knowledge sharing in the Netherlands', *Health Research Policy and Systems*, 10: 33. DOI: 10.1186/1478-4505-10-33.
- Klein, J. (2008) 'Evaluation of Interdisciplinary and Transdisciplinary Research', *American Journal of Preventive Medicine*, 35/2: S116–23.
- Klink, J. et al. (2017) 'Enhancing Interdisciplinary Climate Change Work through Comprehensive Evaluation', *Climate Risk Management*, 15: 109–25.
- Konradt, U. et al. (2016) 'Reflexivity in Teams: A Review and New Perspectives', *The Journal of Psychology*, 150/2: 153–74.
- Langley, G. et al. (2009) *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Lotrecchiano, G. R. (2013) 'A Dynamical Approach toward Understanding Mechanisms of Team Science: Change, Kinship, Tension, and Heritage in a Transdisciplinary Team', *Clinical and Translational Science*, 6/4: 267–78.
- Marshall, M. N. (2014) 'Bridging the Ivory Towers and the Swampy Lowlands; Increasing the Impact of Health Services Research on Quality Improvement', *International Journal for Quality in Health Care*, 26/1: 1–5.
- National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2005) *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press.
- National Research Council. (2015) *Enhancing the Effectiveness of Team Science*. Washington, DC: The National Academies Press.
- Norman, C. D. et al. (2011) 'Evaluating the Science of Discovery in Complex Health Systems', *American Journal of Evaluation*, 32/1: 70–84.
- Okamoto, J. (2015) 'Scientific Collaboration and Team Science: A Social Network Analysis of the Centers for Population Health and Health Disparities', *Translational Behavioral Medicine*, 5/1: 12–23.
- Patton, M. Q. (2000) 'Utilization-Focused Evaluation', in Stufflebeam D. L., Madaus G., and Kellaghan T. (eds) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, pp. 425–38. Boston: Kluwer Academic Publishers.
- Patton, M. Q. (2008) *Utilization-Focused Evaluation*. 4th ed. Los Angeles: SAGE Publications.
- Piggot-Irvine, E., and Zornes, D. (2016) 'Developing a Framework for Research Evaluation in Complex Contexts Such as Action Research', *SAGE Open*, 6/3: 1–15. DOI: 10.1177/2158244016663800.
- Prokopy, L. S. et al. (2015) 'Using a Team Survey to Improve Team Communication for Enhanced Delivery of Agro-Climatic Decision Support Tools', *Agricultural Systems*, 138: 31–7.
- Read, E. K., O'Rourke, M., Hong, G. S., Hanson, P. C., Winslow, L. A., Crowley, S., Brewer, C. A. et al. (2016) 'Building the team for team science', *Ecosphere*, 7/3: e01291. DOI: 10.1002/ecs2.1291.

- Reuveni, Y., and Vashdi, D. R. (2015) 'Innovation in Multidisciplinary Teams: The Moderating Role of Transformational Leadership in the Relationship between Professional Heterogeneity and Shared Mental Models', *European Journal of Work and Organizational Psychology*, 24/5: 678–92.
- Rivera, M. T., Soderstrom, S. B., and Uzzi, B. (2010) 'Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms', *Annual Review of Sociology*, 36/1: 91–115.
- Schippers, M. C., West, M. A., and Dawson, J. F. (2015) 'Team Reflexivity and Innovation: The Moderating Role of Team Context', *Journal of Management*, 41/3: 769–88.
- Siedlok, F., and Hibbert, P. (2014) 'The Organization of Interdisciplinary Research: Modes, Drivers and Barriers', *International Journal of Management Reviews*, 16/2: 194–210.
- Spitters, H. P. E. M. et al. (2017) 'Developing a Policy Game Intervention to Enhance Collaboration in Public Health Policymaking in Three European Countries', *BMC Public Health*, 17: 961.
- Stetler, C. B. et al. (2006) 'The Role of Formative Evaluation in Implementation Research and the QUERI Experience', *Journal of General Internal Medicine*, 21/Suppl 2: S1–8.
- Stokols, D. et al. (2008) 'The Ecology of Team Science: Understanding Contextual Influences on Transdisciplinary Collaboration', *American Journal of Preventive Medicine*, 35/Suppl 2: S96–115.
- Strasser, D. C. et al. (2010) 'Measuring Team Process for Quality Improvement', *Topics in Stroke Rehabilitation*, 17/4: 282–93.
- Trochim, W. M. et al. (2008) 'The Evaluation of Large Research Initiatives: A Participatory Integrative Mixed-Methods Approach', *American Journal of Evaluation*, 29/1: 8–28.
- Tröster, C., Mehra, A., and van Knippenberg, D. (2014) 'Structuring for Team Success: The Interactive Effects of Network Structure and Cultural Diversity on Team Potency and Performance', *Organizational Behavior and Human Decision Processes*, 124/2: 245–55.
- Tsai, C. C., Corley, E. A., and Bozeman, B. (2016) 'Collaboration Experiences Across Scientific Disciplines and Cohorts', *Scientometrics*, 108/2: 505–29.
- Van Noorden, R. (2015) 'Interdisciplinary Research by the Numbers', *Nature*, 525/7569: 306–7.
- Vindrola-Padros, C. et al. (2017) 'The Role of Embedded Research in Quality Improvement: A Narrative Review', *BMJ Quality and Safety*, 26/1: 70–80.
- Vogel, A. L. et al. (2013) 'The Team Science Toolkit', *American Journal of Preventive Medicine*, 45/6: 787–9.
- Willis, C. D. et al. (2017) 'Evaluating the Impact of Applied Prevention Research Centres: Results from a Modified Delphi Approach', *Research Evaluation*, 26/2: 78–90.
- Wu, Y., and Duan, Z. (2015) 'Social Network Analysis of International Scientific Collaboration on Psychiatry Research', *International Journal of Mental Health Systems*, 9/1: 2.
- Yin, R. K. (2009) *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications.