

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Leadership Development in Complex Health Systems: A Qualitative Study
<b>AUTHORS</b>	Curry, Leslie; ayedun, Adeola; Cherlin, Emily; Allen, Nikole; Linnander, Erika

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Angus Ramsay University College London, UK
<b>REVIEW RETURNED</b>	18-Dec-2019

<b>GENERAL COMMENTS</b>	<p>Many thanks for the opportunity to review this revised paper. Overall, it seems to me that the authors have made a great deal of effort to engage with the previous editors' and reviewers' comments: they take each recommendation and question on board and in many cases have addressed them.</p> <p>The paper reads well and presents a clear case that participants had a positive experience of training, which they took forward into their day-to-day leadership practice. I think my main issue is whether enough is said in the manuscript about how this was achieved – without which I am uncertain that readers will gain a great deal of knowledge about how leadership development programmes work. Therefore, I think a couple of the concerns raised by the Editors – that this manuscript might be perceived as PR, and that there is insufficient exploration of negative findings – remain.</p> <p><b>ABSTRACT</b></p> <p>1. This was clearly written, though might need to be updated if the manuscript is amended in any significant way.</p> <p><b>BACKGROUND</b></p> <p>2. I think this section provides a good case for the analysis presented. One thought I had was whether the balance was quite right – that is much more information is provided about the gaps in training programmes themselves than there is about the gaps in knowledge about how training programmes work/provide value. The authors might consider rebalancing this somewhat.</p> <p><b>METHOD</b></p> <p>3. While mentioned in relation to the topic guide probes, it might be useful to emphasise that participants were encouraged to provide negative as well as positive feedback.</p> <p>4. Somewhere – possibly here – the absolute rigour and independence of the team should be made plain. Given that the</p>
-------------------------	--

	<p>programme and several authors are from the same institution, the authors should be absolutely clear in addressing this.</p> <p><b>RESULTS</b></p> <p>5. It is great that participants had such a positive experience of the course. However (in line with the Editors' previous feedback), I would argue that there remains a risk that this piece could be seen as PR. Specifically, the analysis does not – as yet – address the aim to “identify specific structural and curricular design characteristics of leadership development programmes that may promote learnings that sustain over time”. Experiences only tell one part of the story, and exclude elements that readers might be more interested in. I think readers might feel that they are reading about a great course, but not how it achieved these effects. I'll give two examples, but they illustrate a broader point: the authors could do more to describe HOW this successful course and environment have been operationalised – otherwise, readers are left too much with a sense that the course was great, but not how this was achieved.</p> <p>6. Example A: one subsection describes creating a supportive environment/safe space, but very little is said about how this (undoubtedly important) context for learning was actually created.</p> <p>7. Example B: Similarly, the diversity of roles in the cohort is certainly a positive, but can potentially represent a challenge for facilitation (e.g. brokering interaction across organisational or professional hierarchies): did this simply not come up, or were processes in place to address/manage this? At the moment</p> <p>8. There might be also value in exploring any challenges experienced or gaps in the programme and how they were addressed. For instance, did the programme develop over the three years (if so, how/why)?</p> <p><b>DISCUSSION</b></p> <p>9. I'd like to see a bit more of a critical tone. For instance, there are a few references to the Yale Methodology, and it would be good to get some reflection from the authors on whether other e.g. problem-solving strategies (if implemented in such an apparently effective way) could also support similar breakthroughs.</p> <p>10. Under Limitations, the authors state that their sampling/recruitment strategy meant “we did not select only those who had a positive view of the programme”. However, the feedback seems pretty uniformly rosy: what were the criticisms? Further, I'd have thought that social desirability response bias might well have been a risk, no matter what level of response rate.</p> <p><b>MINOR POINTS</b></p> <p>11. There are a few points in the results where the narrative switches between present and past tense, e.g. “Participants DESCRIBE working in complex systems with “difficult landscapes” and DEVELOPED an appreciation for being prepared to address this complexity” (p10). I'd suggest amending to the latter where appropriate.</p> <p>12. missing word: “We resolved differences IN coding by consensus” (p5)</p>
--	---

	13. As noted by the Editors, (Comment 11), participants' previous experience of training is certainly likely to influence their assessment of the programme discussed in this paper. Therefore it might well be worth acknowledging this in the analysis (in the way the authors respond to this comment) and add a supplementary file.
--	---

<b>REVIEWER</b>	Roberta E Sonnino MD FACS FAAP Professor of Pediatric Surgery (Retired) Wayne State University School of Medicine Detroit, MI USA
<b>REVIEW RETURNED</b>	28-Jan-2020

<b>GENERAL COMMENTS</b>	<p>The paper discusses a medium-term evaluation of a Leadership training program in complex health systems. The program has 3 cohorts spanning a 5 year period - therefore the longest follow up period is 4 years from completion of the program. The program reviewed includes residential-style modules, field work components and a specific project for each participant that encourages project-based learning. I concur that these are key elements of a successful leadership program that will allow retention of the skills learned over time.</p> <p>I agree with many of the points the authors make in terms of what competencies and experiences an effective leadership training program should include, such as creativity in strategic problem solving, managing organizational dynamics, interdisciplinary problem-focused leadership development, conflict resolution, as well as a professional network and practical tools, with a level of learning that "sustains and amplifies over time with increased complexity in their work". The life-long support of learning community/peers encountered during these programs is priceless. The concept of a "safe learning community " cannot be over-emphasized - difficult experiences that would remain hidden are often shared and become educational material for all participants in such a setting.</p> <p>Participant observations, i.e. the "results" of this study, are very significant, if not entirely novel. The observation by a participant that "Conducting a field project allowed participants to "very, very swiftly" apply the theory to practice....and prevented the learnings from "dissipating" is one that I have personally seen in other programs. I believe there is enough global experience with this type of practical, experiential learning in varied settings to conclude that the authors and participants are correct about its relevance. Participants also noted that their experience with intergroup dynamics and diverse peers "cemented" relationships. They also noted that stakeholder engagement as well as team activities created a new dynamic for them and made it easier to work with others later in their careers: again, these are important results when one hopes to improve the teams leading complex health systems.</p> <p>Possibly the most interesting comment by a participant is that "the idea that you should devote a much higher proportion of your time to trying to understand what the problem is before you start trying to dream up solutions was a transforming insight for me". If more leaders understood this "aha moment", a lot of problems would reach a prompt and more satisfactory resolution.</p>
-------------------------	--

	<p>There was repeated discussion about how learnings became "embedded" or "ingrained": this is a very important result. The authors conclude that "learnings 'stuck' and grew over time for several reasons: the intentional creation of a diverse learning community, the integration of leading-edge scholarship into a practical 9-month field project to embed new habits, and the curricular balancing of tools and approaches for both strategic thinking and intergroup management." I would add that both the resident-style experiences and the inter-session practical projects most likely also played a significant role.</p> <p>I do have a few comments/questions for the authors:</p> <p>1.No mention was made as to whether participants underwent personality type (such as Meyers-Briggs) and 360 degree assessments before starting the program (with feedback for both at the beginning of the program): I would ask the authors to comment on this, as in my experience, I have found these tools to be eye opening and of great value to aspiring leaders. Again, the "safe" environment plays an important role in processing the feedback received, in ways that cannot be duplicated in other environments.</p> <p>2. While they emphasize that there are few other programs with the same scope of the one studied here, the authors do NOT mention, review or compare those programs. Among these I would mention the ELAM (Executive Leadership in Academic Medicine) program, now in its 25th year, and therefore with a long follow-up, and a great deal of outcomes research. ELAM meets and exceeds the criteria mentioned in this paper except one: it only includes women, as it was devised over 25 years ago to increase the ranks of women leaders in the academic health community. I believe that reviewing some of the ELAM research literature would in fact validate the findings of the present study. Another program worth noting is the University Of Minnesota Medical School's "Emerging Physician Leaders Program (EPLP)". This is a three-year course designed to develop a community of physician leaders for the Medical School. While it is offered only to mid-level faculty of the school, many of the methods and topics validate what is discussed in the current manuscript.</p> <p>In summary, I think this is an excellent manuscript. A bit more detail on the individual course curriculum and a comparison with other comprehensive programs that apply similar strategies, would enhance the importance of the authors' findings. I would encourage the authors to add some comments in that regard.</p>
--	--

<b>REVIEWER</b>	Martin McNamara University College Dublin, Ireland
<b>REVIEW RETURNED</b>	30-Jan-2020

<b>GENERAL COMMENTS</b>	<p>1. The title does not reflect the objectives, focus and substance of the paper that is concerned with participants' experiences of a specific programme rather than leadership development as such.</p> <p>2. Participants' experiences are mainly positive. The interview discussion guide in Box 1 shows no evidence of questions or prompts designed to elicit information on negative experiences or aspects of the programme; e.g., What did not work so well for</p>
-------------------------	---

	<p>you? Was there anything you would like more of? Less of? What aspects do you find challenging to apply in your current practice and context? Would you should others take the programme? Why should they not?</p> <p>3. Participants found certain aspects of the programme to be valuable, but we are told very little about these aspects. For Theme 1, we are told very little about what the newly-generated scientific evidence is and why it was regarded as so compelling. There is scant detail on the project methodology employed for the field work element. There is also very little detail on the content (concepts, theories, models, frameworks and so on) that was considered to be especially relevant. For Theme 2, we have very little information on the underpinning principles governing network formation how they were facilitated. Nor are the practical tools elaborated. For Theme 3, very little data is presented to support the theme. Also for the most recent cohort, very little can be inferred from its participants' responses concerning enduring impact and sustainability, given the recency of completion.</p> <p>4. In short, it is hard to discern the contribution that this paper makes to the literature on the design and delivery of leadership development programmes. Certain 'core content' areas delivered over two modules, combined with a field-based project elicited mainly positive feedback from participants at three time sampling points. However, we know very little about what ideas and tools added what specific value for whom to what extent, why and how.</p> <p>5. A possible solution to the problems caused by compromising depth for breadth in this paper might be to devote a paper to each theme and consider what specific knowledge (theories, models, concepts, frameworks), tools and methodologies proved so compelling and valuable for participants' practice, and for their personal and professional development. However, I acknowledge that there may be commercial reasons why the authors' do not wish to share this level of detail of curriculum design and delivery, although the references are suggestive. The problem is that without adequate elaboration of key concepts such as 'strategic problem solving' and 'adaptive leadership', it is very hard to appreciate exactly what facets participants' found so valuable and why.</p>
--	---

## VERSION 1 – AUTHOR RESPONSE

### **Reviewer 1 Comments and Responses**

#### **Comment 1**

This paper reads well and presents a clear case that participants had a positive experience of training, which they took forward into their day-to-day leadership practice. I think my main issue is whether enough is said in the manuscript about how this was achieved – without which I am uncertain that readers will gain a great deal of knowledge about how leadership development programs work.

#### **Response:**

We very much appreciate the reviewer's reflection that the manuscript did not sufficiently convey how the course content and learning environment allowed participants to take the learnings forward to their day-to-day leadership practice. As a major substantive revision, we expanded Box 1 to include comprehensive details on specific structural and design features of the programme (with related citations) to describe how the programme works. Revised Box 1, newly titled, "Description of Programme Objectives, Structural and Curricular Features" appears on page 4 of the manuscript, as follows:

### **Box 1**

#### **Description of Programme Objectives, Structural and Curricular Features**

<p><b>Objectives:</b></p> <ul style="list-style-type: none"> <li>• Strengthen leadership capability among teams to solve complex challenges in health and social care</li> <li>• Foster effective relationship management across diverse groups and system-wide boundaries</li> <li>• Provide the analytical and behavioral science tools needed to deliver robust strategic change within their systems</li> <li>• Explore the international context of health and social care through a comparative lens</li> </ul>
<p><b>Partner roles:</b></p> <ul style="list-style-type: none"> <li>• UK partner manages the participant application and selection process, provides psychometric testing and fitness services, identifies and arranges speakers from the UK health and social care system, manages programme logistics in UK and travel to the US, and provides periodic check-ins to participants as they complete the field project</li> <li>• US partner manages and delivers curricular content, identifies and arranges US site visits and speakers, manages programme logistics in US, provides feedback to participants at field project milestones, and certifies completion</li> </ul>
<p><b>Overall design:</b></p> <ul style="list-style-type: none"> <li>• 2 Residential 6-day retreat modules (first in UK, second in US)</li> <li>• 9-month field project to address a complex adaptive challenge in health and social care</li> <li>• Certificate of completion issued by Yale School of Public Health</li> </ul>
<p><b>Participant composition:</b></p> <ul style="list-style-type: none"> <li>• Individuals (cohorts 2013/14 and 2015/16; teams of 3 (beginning in 2017/18)</li> <li>• Diverse individuals (e.g. clinical and non-clinical backgrounds, gender and other protected characteristics)</li> <li>• Cross organizational and sectoral representation (health, social care, local government, most recent cohort)</li> <li>• Sufficient level of seniority within their organization to effect change</li> <li>• Endorsement from senior leader for time in the programme and support for the field project</li> </ul>

<p><b>Curricular content</b></p> <p><i>Didactics and table top exercises</i></p> <ul style="list-style-type: none"> <li>• Working across boundaries(Carlise, 2004;Santos, 2005) and in hierarchy(Jaques 1990;Ronay 2012)</li> <li>• Leadership and followership(Berg, 1990)</li> <li>• Representational group theory(Alderfer, 1987; Berg, 2005;Curry 2012)</li> <li>• Psychological safety(Nembhard, 2006)</li> <li>• Levels of analysis(Alderfer, 2011; Berg, 2015)</li> </ul> <p><i>Keynote lectures</i></p> <ul style="list-style-type: none"> <li>• UK speakers on national strategic priorities and current reforms</li> <li>• Breaking science from the Yale team’s research on leadership, management, culture and organizational performance</li> <li>• Site visits and roundtables with practitioners/experts on US health and social care</li> </ul> <p><i>Strategic problem-solving methodology</i>(Bradley, 2019;Curry, 2013)</p> <ul style="list-style-type: none"> <li>• Define the problem and objective</li> <li>• Identify and prioritize root causes</li> <li>• Generate and pursue strategic solutions</li> <li>• Measure progress and impact</li> </ul> <p><i>Individual Psychometric evaluations (optional)</i></p> <ul style="list-style-type: none"> <li>• 15FQ+;(PSYTECH International) Watson-Glaser Critical Thinking Appraisal (UK);(Watson, 1952) Myers-Briggs Type Indicator (MBTI)(Myers-Brigg Company); StressScan Indicator(Lazarus, 1984)</li> <li>• Administered via online survey, followed by personal feedback sessions with a UK-based organizational development consultant during the first programme module</li> </ul>
<p><b>Field project:</b></p> <ul style="list-style-type: none"> <li>• Participants are supported to: <ul style="list-style-type: none"> <li>○ Identify a problem of strategic importance to the organization or system</li> <li>○ Convene a guiding coalition(Bradley, 2018; Kotter, 2001) of diverse stakeholders that can understand and address the problem</li> <li>○ Facilitate the guiding coalition through the strategic problem solving process, integrating new perspectives on organizational dynamics</li> </ul> </li> <li>• Participants produce deliverables for faculty feedback: <ul style="list-style-type: none"> <li>○ Problem statement, objective, coalition members, and sponsor sign-off</li> <li>○ 3-page written summary of project progress and leadership reflections</li> <li>○ Oral presentation to faculty and peers during module 2</li> </ul> </li> </ul>
<p><b>Supports for participant wellbeing:</b></p>

- Optional daily group exercise class and individual fitness/nutrition consultations
- Alderfer CP. An intergroup perspective on group dynamics. In: Lorsch J, ed. Handbook of organizational behavior. Englewood Cliffs, NJ: Prentice Hall, 1987:190-222.
- Alderfer CP. *The practice of organizational diagnosis*. New York: Oxford University Press, 2011.
- Berg DN. Levels of analysis: diagnosis & intervention. <http://csms.org/wp-content/uploads/2015/04/Levels-Organizational-Analysis.pdf> (accessed December 7, 2016).
- Berg DN. Resurrecting the muse: followership in organizations. In: Klein EB, Gabelnick F, Herr P, eds. The psychodynamics of leadership. Madison. CT: Psychosocial Press, 1990:27-52.
- Berg DN. Senior executive teams: not what you think. *Consult Psychol J: Pract and Res* 2005.
- Bradley EH, Brewster AL, McNatt Z, et al. How guiding coalitions promote positive culture change in hospitals: a longitudinal mixed methods interventional study. *BMJ Qual Saf* 2018;**27**(3):218-25.
- Bradley EH, Nembhard I, Taylor L, et al. Leadership and management: a framework for action. In: Burns L, Weiner B, eds. Shortell and Kaluzny's Health Care Management: Organization Design and Behavior. New York: Delmar Cengage Learning, 2019:32-55.
- Carlile PR. Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organ Sci* 2004;**15**(5):555-68.
- Curry LA, Krumholz HM, O'Cathain A, et al. Mixed methods in biomedical and health services research. *Circ Cardiovasc Qual Outcomes* 2013;**6**(1):119-23.
- Curry LA, O'Cathain A, Clark VLP, et al. The Role of Group Dynamics in Mixed Methods Health Sciences Research Teams. *J Mix Method Res* 2012;**6**(1):5-20.
- Jaques E. In praise of hierarchy. *Harv Bus Rev* 1990;**68**(1):127-33.
- Kotter JP. What leaders really do. *Harv Bus Rev* 2001;**79**(11):85+.
- Lazarus RS, Folkman S. *Stress, appraisal and coping*. New York: Springer, 1984.
- Nembhard IM, Edmondson AC. Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *J Organ Behav* 2006;**27**(7):941-66.
- PSYTECH International. Fifteen Factor Questionnaire Plus. <https://www.psytech.com/Assessments/FifteenFactorQuestionnairePlus> (accessed February 6, 2020).
- Ronay R, Greenaway K, Anicich EM, et al. The path to glory is paved with hierarchy: when hierarchical differentiation increases group effectiveness. *Psychol Sci* 2012;**23**(6):669-77.
- Santos FA, Eisenhardt KA. Organizational boundaries and theories of organization. *Organ Sci* 2005;**16**(5):491-508.
- The Myers-Briggs Company. Myers-Briggs Type Indicator®. <https://www.cpp.com/products/mbti/index.aspx> (accessed February 6, 2020).
- Watson G, Glaser EM. *Watson-Glaser Critical Thinking Appraisal*. Yonkers, New York: World Book Company, 1952.

I think a couple of the concerns raised by the Editors - that this manuscript might be perceived as PR, and that there is insufficient exploration of negative findings, remain.

**Response:**

We thank the editors and reviewers for raising these concerns. We have made major revisions to the manuscript in response, as follows.

**Concern the manuscript might be perceived as PR:** We have been mindful of this concern throughout the development of the manuscript. We believe that several revisions, particularly in response to these most current reviews, minimize the potential for this perception. First, as detailed in the next paragraph, we have made substantive additions to reflect additional negative findings in a comprehensive new Supplemental Table and to reflect additional negative findings both in the text as feasible given space constraints. Second, as detailed below in response to Reviewer 1, comment 8, we have presented a more critical tone to the discussion section by further developing the study limitations as well as clarifying that the “Yale methodology” (participants’ words) is simply one of a number of problem-solving methods and stated that others may be just as effective. Finally, while we include the specific name of the programme and partners as we believe readers will be interested in the context for the study, we defer to the editors whether this should be omitted.

**Insufficient exploration of negative findings:** We very much appreciate this feedback. We used several established methodological techniques to explore negative findings, including multiple specific probes in the discussion guide, and special attention to and analysis of critical comments in the data. However, we agree with the reviewer that the manuscript did not adequately convey those findings and have made substantive additions to the manuscript accordingly. First, we compiled all negative or critical quotes in a new Supplemental Table 2 titled, “Negative Experiences and Reflections Reported by Participants” (please see below). While some quotes have been edited for length, they represent the complete set of all comments with a negative or critical tone in the dataset. We would also note that this is a very small portion of the overall dataset (300 pages of transcripts). We believe this new table is a useful addition and thank the editors and reviewers for this recommendation. Readers are directed to the table on page 9 of the manuscript, as follows:

Some participant feedback included critical or negative comments related to the following aspects of the programme: networking opportunities, course content, participant selection and cost. All comments are reported in Supplemental Table 2.

**Supplemental Table 2**

**Negative Experiences and Reflections Reported by Programme Participants**

*Note: Interviewers probed for negative feedback with questions such as: What did not work so well? What was not as worthwhile? What was not as relevant? This table includes all pieces of negative feedback in the dataset. Some quotations were edited for length.*

Theme	Quote
Lack of networking opportunities for alumni across cohorts	<p><i>“I haven’t had any lasting professional relationships with other delegates on the course.” (Interview 1 Male, Board Member)</i></p> <p><i>“I haven’t really kept in touch with many people post-program. I think that that’s a bit of a shame, really, ‘cause I really got on with the people whilst on the program.” (Interview 4 Male, Board Member)</i></p>

	<p><i>"I felt in the first half of the program that was delivered over here, I got to know the people from my group who I didn't know all that well before. I didn't really get to know the other people very well at all. I think that's 'cause the classroom teaching format wasn't particularly conducive to that." (Interview 15 Male, Board Member)</i></p> <p><i>"What I found disappointing, which I have not seen in the academic NRHR leadership world is that there was basically no communication with other groups and no way to facilitate people who've gone through the same project to see whether there would be joint possibilities to work on strategies." (Interview 16 Female, Senior Manager)</i></p> <p><i>"There are a whole host of people in the system that have been on the Yale course. The bond is not there just because you've been on the Yale course. I know there's been some sort of [alumnae reunion] sessions, and maybe people that were invited to those sessions coming back would have benefited from that, but I wasn't invited to those sessions." (Interview 24 Female, Board Member)</i></p> <p><i>"I would have really appreciated some sort of organized meet up six months later, then a year. Just a way to keep some of that energy, that motivation that you always get from a good course. How do you then keep some of that and maintain it going forwards. That would be the only bit for me...It would be really useful to keep some of that going. (Interview 21 Male, Clinician, Board Member)</i></p> <p><i>"On the slightly less positive is that I know in my region there are now a number of people who've done this program, some of which are very senior and have leadership roles in the STP. I've been disappointed that I don't see the impacts of the program in terms of what we as a team, as a group of leaders in the STP, are doing." (Interview 12 Female, Board member)</i></p>
<p>Gaps between theory and practice</p>	<p><i>"I worked very hard on one project, for example, and I've tried to fit all the learning that I've had from the program in to practice... The program needs to be backed up by me practicing all the time now, but it doesn't equip you with everything that you need." (Interview 2 Female, Senior Manager)</i></p> <p><i>"A large amount of the course was working on root cause analysis and developing solutions by a fairly didactic system which I haven't found that useful... The projects that we were made to do in a very kind of prescriptive way, whereas we tend to find in real life, it's not done in such a linear fashion." (Interview 25 Male, Board Member)</i></p>
<p>Gaps in course content</p>	<p><i>"I struggled a little bit with the amount of time we spent on some of the tools and working up almost like a pretend project to use the tools... the one session that didn't seem to work for me was the expert roundtable sessions, I didn't get as much from those conversations as I did from the other sessions." (Interview 006 Male, Senior Manager)</i></p> <p><i>"One of the things that would be useful would've been to say, well, this is how we recommend you approach this system-wide project. Let's try it out. This is how you get others to understand this methodology. Maybe we did do that, and it just didn't land very well with me." (Interview 12 Female, Board Member)</i></p> <p><i>"The first half was in England. There were some talks by invited speakers who I had thought would be informative. In fact, I think they were really very uninformative. I would've preferred to perhaps have that time just to network with the other people in the group, find out what they've been doing and so on." (Interview 15 Male, Board Member)</i></p>

<p>Participant Selection</p>	<p><i>“If you were very junior you might feel like you’re out of place a little bit.” (Interview 2 Female, Senior Manager)</i></p> <p><i>“I’ve never really done a course as part of a group. It’s always been on an individual basis. I don’t think I’d really thought through to start with how beneficial that might be or risky it might be. I guess the risk with it is if you don’t know your group terribly well and you don’t all get on, that could be quite tricky.” (Interview 9 Female, Senior Manager)</i></p> <p><i>“I think the course would have worked better if there’d have been a group of three of us from our area. I think now that is a requisite that you need to actually attend as a group rather than an individual...I think I was just in that period where you could get in as an individual. I don’t think it works as well.... It’s difficult, ‘cause the groups that I’m in seem to work, and yet they’ve not all been on the course. Did I need it to get that? I don’t know...I really did enjoy the course, but I’m not sure other people will. If I had another two weeks, would I do it again? Yes, I probably would, but I know I’d do it differently. I would go with a team and would actually plan what we were gonna do before we went to the first week.” (Interview 25 Male, Board Member)</i></p> <p><i>“I guess what I was frustrated at year on year is seeing that the course was still largely promoted to people in the NHS because increasingly, it’s a system approach. In my mind, it needed to be more widely promoted to the local authority and actually doing the course with someone say from the local authority, from commissioning, from a number of providers, from primary care.” (Interview 24 Female, Board Member)</i></p>
<p>Programme cost</p>	<p><i>“If one of my staff approached me and said, “I would like to go to Yale. It will cost us x” I would say that there are lots of leadership courses here that I could pay for you to go on and it would be a lot cheaper. I’m afraid that’s just my perspective as the budget holder. I’m not sure I would feel comfortable in allowing that much money to be spent...If I got the chance to go again, I absolutely would. I just wanted to give you a rounded response. Even if I had to approach my boss and I asked her a set amount of money, I think there would be questions.” (Interview 20 Male, National Level)</i></p> <p><i>“The course was not cheap...It wasn’t about the quality of the course. I’m not necessarily saying it was expensive for what you got. I’m just saying that there was no way my organization was going to allow me the time out to do it again.” (Interview 23 Female, Board Member)</i></p> <p><i>“It’s not a cheap course to go, especially going abroad for a week. That’s not something that you would normally fund within the health services. You just can’t fund it.” (Interview 24 Female, Board Member)</i></p>
<p>Change in perspective</p>	<p><i>“I think it was a sort of one-hit thing. I went on the program, I learned what I learned, I gained the insights I gained from the program, and I’ve applied those over time, but I haven’t changed how I look back on the course.” (Interview 1 Male, Board Member)</i></p>

	<p><i>“There’s nothing that is dramatic that has changed in the way that I work...It’s all about every time you do one [program], it’s about some incremental change and additional thought process that you go through. No, I thought it was good when I left. I haven’t had an epiphany where I thought, “Wow, this has changed things completely since then.” (Interview 20 Male, National Level)</i></p> <p><i>“With distance, there’s elements of it that I have probably forgotten. There’s elements of it that I couldn’t maybe articulate. Where it might have changed some of my behaviors, I wouldn’t be able to tell you exactly.” (Interview 14 Female, Senior Manager)</i></p>
<p>Other negative experiences or views</p>	<p><i>“The insights from the course really shape how I would run a seminar for aspiring healthcare leaders...I think if you went back to my hospital and said “Did we really get value for money out of the programme, they might well say, ‘Well, not really. He moved on pretty quickly, and he’s not involved in senior-level healthcare leadership himself anymore. He might be supporting other people in those roles, but he is not doing it himself.” (Interview 1 Male, Board Member)</i></p> <p><i>“I’m gonna be honest here. I loved going to Yale, and I loved the opportunity to spend 24 hours in New York beforehand, but I’m not quite certain what being in Yale added to the learning experience. Yale is a very special place, but I wasn’t quite sure, from my perspective, as to what it added to my learning from the course.” (Interview 22 Female, Board Member)</i></p> <p><i>“I didn’t find the visits to American healthcare institutions particularly helpful. It was interesting, but it didn’t change my thinking in any way. I think we went to a lovely American hospice, and it was interesting to see. They’ve clearly got a lot of resources, and a lot of thought and care had gone into the practice, but there wasn’t much that I could do with that information in my world.” (Interview 1 Male, Board Member)</i></p> <p><i>“A minor, minor thing. It by no means reflects on the value of the program, but on the very, very first day...the very first thing was an introduction. Everybody around—what’s their name, where they were born, what difficult challenges they found growing up and how did they overcome it. I’m quite an introvert, and it doesn’t come easy to me talking about this level of personal detail, and perhaps that was a little bit too intense on the very first day when we didn’t know anybody at all.” (Interview 2 Female, Senior Manager)</i></p> <p><i>“I think what we are missing is the impact this program has then actually had....What is the impact that you think this program had? You did not ask me about any specific projects. You did not ask me, what was the actual impact on the NHS?” (Interview 16 Female, Senior Manager)</i></p> <p><i>“I thought some of his stuff was really good, but I think most people thought he [a speaker] was completely fantastic. It was just I thought there was almost a degree of hero worship around him sometimes.” (Interview 22 Female, Board Member)</i></p> <p><i>“Comparing the U.S. and the British healthcare systems. Britain is different from the U.S., and we’re not gonna suddenly become the U.S...It’s interesting to find out what’s going on in France or Syria or wherever, but it’s not something that has a massive impact on one’s life.” (Interview 1 Male, Board Member)</i></p>

	<p><i>"I found that there were some really interesting concepts I wanted to cling onto, but we'd move on to something else. I would then be struggling to remember what those were and really work out. There's working up how what you've just learned that impacts your role going forward." (Interview 20 Male; National Level)</i></p>
--	--

Second, we made the following revision to convey the divergent views of several participants on certain points:

Page 13: First, most, but not all, participants asserted the programme cultivated a robust social network that has provided them with "*common language and common understanding*," allowing them to increase their impact as they address problems within their systems.

Page 15: Participants reflected that programme learnings '*stuck*' with them ("*...let people have their groups' is something that possibly every day in my working career since then has stuck with me*"), although ~~one~~ some shared learnings "had worn off a bit".

Pages 15-16: Of note, one participant described the changes as "incremental" and not as readily perceivable and for another, the relevancy of concepts was not immediately apparent as they wondered "how valuable any of it had been."

### Comment 3

BACKGROUND: I think this section provides a good case for the analysis presented. One though I had was whether the balance was quite right – that is much more information is provided about the gaps in training programmes themselves than there is about gaps in knowledge about how training programmes work/provide value. The authors might consider rebalancing this somewhat.

#### Response:

We appreciate the suggestion to somewhat rebalance the introduction to highlight gaps in knowledge about how health leadership training programmes work. We identified a recent comprehensive systematic review that determined there is very little evidence about best practices in a wide range of programme elements. Mindful of space, we added the following content and citation to the introduction on page 4 of the manuscript:

A 2014 systematic review of 250 health leadership education programs concluded that, due to high variability in programs and limited evaluation data, it was not possible to identify best practices in terms of learning context, content, activities, and delivery mechanisms. (Careau et al., 2014)

### Comment 4

METHOD: While mentioned in relation to the topic guide probes, it might be useful to emphasize that participants were encouraged to provide negative as well as positive feedback.

#### Response:

We appreciate this suggestion, as well as reviewer 3's comment regarding the need to clarify that negative feedback was encouraged from participants. We used multiple types of probes to this end, as currently briefly noted with citations in the manuscript on page 7 ("*Probes were used to generate 'thick descriptions' of their experiences, as well as to elicit both positive and negative views (Creswell & Miller, 2000; Denzin, 1989)*") However, we appreciate that clarity on this point is important, and have added the following both on page 7 of the manuscript and to the discussion guide in Box 2, in order to provide specific examples of probes:

Probes: Participants were encouraged to share negative views through probes such as: What did not work so well? What was not as worthwhile? What was not as relevant?

### Comment 5

METHOD: Somewhere – possibly here, the absolute rigour and independence of the team should be made plain. Given that the programme and several authors are from the same institution, the authors should be absolutely clear in addressing this.

#### Response:

We thank the reviewer for this important observation and recommendation. We have added the following additional clarification regarding rigor and independence on page 7 of the manuscript:

All members of the research team had some experience with the programme, consistent with the principle of ‘the evaluator as instrument’ (Guba & Lincoln, 1981) to generate novel insights and situate the findings in what is known We intentionally assigned team roles to maximize the generativity of data collection efforts and to capitalize on experts’ skills in the analysis phase (where all data were anonymous). Interviews were conducted by two research associates (AA, NA) with experience in developing rapport while preserving social and intellectual distance. (Hammersley & Atkinson, 1983) We anticipated there would be few, if any, power dynamics in the conversation given the interviewers were research assistants who have no power to impact participant’s careers or professional development. Because participants have nothing to gain from reporting positive experiences, or to lose from reporting negative experiences, we anticipated that fear of reprisal or social desirability bias would be minimal. In addition, we believe alumni desire to provide constructive criticism was likely to counterbalance social desirability bias toward positive reflections; nevertheless, we note social desirability as a possible limitation of the study.

In addition, we added an explicit limitation regarding the potential for bias given the relationship between the authors and the programme, on page 18 of the manuscript as follows:

Second, while there are benefits of having the evaluation team be familiar with the programme in terms of the quality and rigor of data collection and interpretation (see for example Guba’s Evaluator as Instrument),(Guba & Lincoln, 1981) we recognize the potential risk to objectivity. Therefore, to ensure transparency, we report all negative comments in the dataset in Supplemental Table 2.

### Comment 6

RESULTS: It is great that participants had such a positive experience of the course. However (in line with the Editors’ previous feedback), I would argue there remains a risk that this piece could be seen as PR. Specifically, the analysis does not – as yet- address the aim to “identify structural and curricular design characteristics of leadership development programmes that sustain over time”. Experiences only tell one part of the story, and exclude elements that readers might be more interested in. I think readers might feel that they are reading about a great course, but not how to achieve these effects. I’ll give two examples, but they illustrate a broader point: the authors could do more to describe HOW this successful course and environment was operationalized – otherwise, readers are left too much with a sense that the course was great, but not who this was achieved. EXAMPLE A: One subsection describes creating a supportive environment/safe space, but very little is said about how this (undoubtedly important) context for learning was created. EXAMPLE B: Similarly, the diversity in roles in the cohort is certainly positive, but can potentially represent a challenge for facilitation (e.g., brokering interaction across organizational or professional hierarchies): did this simply not come up, or were processes in place to address/manage this?

## Response

We very much appreciate the reviewer's observation that the readers might be interested in learning more about how the course content and learning environment was operationalized. The two examples provided are very helpful, and issues related to supportive environments and working across hierarchy are intentionally addressed in both the design and content of the programme. As noted above, we have made a major substantive revision, creating a new Box titled, "Description of Programme Objectives, Structural and Curricular Features". The Box presents details on specific structural and design characteristics, with related citations, in order to convey how the programme is operationalized.

## Comment 7

There might be value in exploring any challenges experienced or gaps in the program and how they are addressed. For instance, did the program develop over the three years (if so, how/why)?

### Response:

We thank the reviewer for this important question. To clarify, the three cohorts sampled reflect 6 years of programme experience (2013-2018). The programme has evolved over the 12 years since its inception, to ensure it is adapted to be responsive to priorities and developments in the U.K. For example, as Clinical Commissioning Groups began in 2012 participant selection evolved to focus on clinical leadership. With the rapid transition to STPs and integrated care systems, participant selection changed from individual executives from within the NHS to teams comprised of representatives from the NHS, social care and local government. And with increasing attention to the national need for leadership capacity building, the programme evolved to be housed within the NHS workforce development portfolio. We would be happy to provide more detail about these developments yet are mindful of space considerations and defer to the editor for guidance if this content would be helpful to readers. The study setting description currently reads as follows, on page 5 of the manuscript.

The study was conducted in the context of the Yale Strategic System Leadership Programme for Sustainability and Transformation Partnership Footprints, launched in 2007 and initially underwritten by South Essex Partnership NHS Foundation Trust with participants funded by their organisation or system; subsequent funders include Health Education England and the National Leadership Academy. Participants include senior clinical and non-clinical system leaders in the NHS (including providers, commissioners, and executives) and, beginning in 2017, teams of three from 10 Sustainability and Transformation Partnerships (STPs) including a senior clinical lead, managerial lead and another sector (e.g. local authority, wider public sector, 3<sup>rd</sup> sector).

## Comment 8

Discussion: I'd like to see a bit more of a critical tone. For instance, there are few references in the Yale Methodology, and it would be good to get some more reflection from the authors on whether other e.g., problem-solving strategies (if implemented in such an apparently effective way) could also support similar breakthroughs.

### Response:

We agree with the reviewer about the need for a bit more of a critical tone and have made several revisions accordingly. First, we completely agree with the reviewer that many other problem-solving methodologies, if implemented in an effective way, can support similar breakthroughs. We believe the critical differentiator is tending to both tasks (applying a problem-solving method) AND relationships (supporting the organizational dynamics associated with the work). We have added the following acknowledgement of other approaches and clarification to the discussion on pages 17-18 of the manuscript:

Of note, although participants consistently described the value of the “Yale methodology” (their term), we do not believe it to be a proprietary model. Their descriptions highlighted learnings related to development of shared understanding of the problem through stakeholder engagement and root cause analysis process that incorporates multiple methods and perspectives. We hypothesize that many other step-wise approaches to systematically understanding and addressing complex problems, if implemented in an effective way, would support similar breakthroughs.

In addition, we made edits to delete reference to the “Yale methodology” on page 13 of the manuscript as follows:

Practical, concrete methods such as the ~~“Yale methodology”~~ of strategic problem solving spurred new ways of thinking.

And on page 14 of the manuscript as follows:

*What has been very interesting is beginning to apply the ~~Yale~~ problem-solving methodology to the problems that we’ve had...*

### Comment 9

Under limitations, the authors state that their sampling/recruitment strategy meant “we did not select only those who had a positive view of the programme”. However, the feedback seems uniformly rosy; what were the criticisms? Further, I’d have through that social desirability response bias might have been a risk, no matter what level of response rate.

#### Response:

We thank the reviewer for this feedback and have accordingly revised the limitations section in the following ways to address criticisms and social desirability response bias:

**Criticisms of the programme:** As noted above in response to comment 1, we compiled all negative quotes in a new supplemental table titled, “Negative Experiences and Reflections Reported by Participants.” While some quotes have been edited for length, they represent the complete set of all comments with a negative or critical tone in the dataset. We believe this new table is a useful addition and thank the editors and reviewers for this recommendation. Readers are directed to the table on page 8, as follows:

Some participant feedback included critical or negative comments related to the following aspects of the programme: networking opportunities, course content, participant selection and cost. These comments are reported in Supplemental Table 2.

And again on page 18 of the manuscript, as follows:

Second, while there are benefits of having the evaluation team familiar with the programme in terms of the quality and rigor of data collection and analysis (see Guba’s Evaluator as Instrument), (Guba & Lincoln, 1981) we recognize the potential risk to objectivity; hence, to ensure transparency we report all negative experience and comments in Supplemental Table 2.

**Social desirability risk:** we agree with the reviewer that response rate is not the only factor that might influence social desirability response bias. We have revised the text on pages 18-19 of the manuscript as follows:

Several limitations must be noted. First, social desirability response bias (Sudman, Bradburn, & Schwarz) may have occurred. However, we used established

techniques (Creswell & Miller, 2000; Denzin, 1989) to encourage participants to share both positive and negative experiences. In addition, the (Patton, 2002)randomization of the complete sample list and low refusal rate reduced the likelihood that we did not select only those who had a positive view of the programme. Although the refusal rate was low, it is possible that those who refused or missed the interview had a more negative or neutral experience with the programme. Second, while there are benefits of having the evaluation team familiar with the programme in terms of the quality and rigor of data collection and analysis (see Guba's Evaluator as Instrument),(Guba & Lincoln, 1981) we recognize the potential risk to objectivity; hence, to ensure transparency we report all negative experience and comments in Supplemental Table 2.

#### **Comment 10**

There are a few points in the results where the narrative switches between present and past tense, e.g., "Participants DESCRIBE working in complex systems with 'difficult landscapes' and DEVELOPED an appreciation for being prepared to address this complexity" (P. 10). I'd suggest amending to the latter where appropriate.

#### **Response:**

We thank the reviewer for the careful read and suggestion. We have reviewed the full manuscript closely and believe we have made all necessary corrections to these errors, noted in the marked copy of the document.

#### **Comment 11**

Missing word: "We resolved differences IN coding by consensus (P.5).

#### **Response:**

Again, we thank the reviewer. We apologize for the typographical error and have made this correction.

### **Reviewer 2 Comments and Responses**

#### **Comment 1**

Possibly the most interesting comment by a participant is "the idea that you should devote a much higher proportion of your time to understand what the problem is before you start to dream up solutions was a transforming insight for me". If more leaders understood this "aha" moment, a lot of problems would reach a prompt and more satisfactory resolution. There was repeated discussion about how learnings became "embedded" or "ingrained": this was a very important result. The authors conclude that "learnings stuck" and grew over time for several reasons – the intentional creation of a diverse learning community, the integration of leading-edge scholarship into a practical 9-month field project to embed new habits, and the curricular balancing of tools and approaches for both strategic thinking and intergroup management. I would add that both the resident-style experiences and the inter-session practical projects most likely played a significant role.

#### **Response:**

We thank the reviewer for these observations and appreciate the suggestion to add the residential experiences and inter-session practical projects to the hypothesized reasons that learnings stuck. The inter-session practical field projects were already noted in the manuscript, but we added reference to the residential-style experiences, revising the manuscript on page 18 as follows:

We hypothesize learnings '*stuck*' and grew over time for several reasons: the intentional creation of a diverse learning community through residential-style experiences, the integration of leading-edge scholarship into a practical 9-month field

project to embed new habits, and the curricular balancing of tools and approaches for both strategic thinking and management of organizational dynamics (Berg, 1979).

## Comment 2

No mention was made as to whether participants underwent personality type (such as Meyers-Briggs) and 360-degree assessments before starting the program (with feedback for both at the beginning of the program). I would ask the authors to comment on this, as in my experience, I have found these tools to be eye-opening and of great value to aspiring leaders. Again, the “safe” environment plays important role in processing the feedback received, in ways that cannot duplicated in other environments.

### Response:

We appreciate this important observation about these professional development approaches and should have noted they are part of the programme. Participants had the option to participate in several individual-level psychometric assessments, including 15FQ+; Watson-Glaser Critical Thinking Appraisal (UK); Myers-Briggs Type Indicator (MBTI); StressScan Indicator. These were administered via online survey, followed by personal feedback sessions with a UK-based organizational development consultant during the first programme module. We have included these assessments in the new Box 1 “Description of Programme Objectives, Structural and Curricular Features” as included above. Although a number of leadership programmes use these personal assessments as a jumping off point for reflection and development, our study participants did not identify these components as a valuable part of the experience and they were therefore not included in the most salient themes. We defer to the editors as to whether additional description of this phenomenon is warranted.

## Comment 3

While they emphasize that there are other programs with the same scope of the one studied here, the authors do NOT mention, review or compare those programs. Among these I would mention the ELAM (Executive Leadership in Academic Medicine) program, now in its 25<sup>th</sup> year, and therefore with a long follow-up, and a great deal of outcomes research. ELAM meets and exceeds the criteria mentioned in this paper except one: it only includes women, as it was devised 25 years ago to increase the ranks of women leaders in the academic health community. I believe reviewing some of the ELAM research literature would in fact validate some of the findings of this present study. Another program worth noting is the University of Minnesota Medical School’s “Emerging Physician Leaders Program (EPLP)”. This is a three-year course designed to develop a community of physician leaders for the Medical School. While it is offered only to mid-level faculty of the school, many of the methods and topics validate what is discussed in the current manuscript.

### Response:

We thank the reviewer for this feedback and are very encouraged to hear that programs such as ELAM and EPLP validate findings in this study. We appreciate the reviewer’s suggestion to review and compare similar programs; unfortunately, space constraints preclude an adequately comprehensive review of other programs. However, we agree that ELAM has demonstrated positive enduring changes through longitudinal evaluations and has also emphasized the value of working in diverse teams. We have added references to ELAM evaluation findings in two places in the discussion on page 17:

The best leadership development programmes produce enduring changes in employee behavior (Dannels et al., 2008; Kirkpatrick & Kirkpatrick, 2005); participants reported leadership approaches and behaviors that not only sustained but in fact amplified over time.

And again on page 18:

*Leadership* development (as compared to *leader* development) requires cultivating capacity to address the relational aspects of work, and to work effectively in diverse teams. (Berg, 1979; Careau et al., 2014; Dannels et al., 2008; Day, 2000)

#### Comment 4

In summary, I think this is an excellent manuscript. A bit more detail on the individual course curriculum and a comparison with other comprehensive programs that apply similar strategies, would enhance the importance of the authors' findings. I would encourage the authors to add some comments in that regard.

#### Response:

We appreciate the reviewer's encouragement and agree that more detail on the individual course curriculum was needed. As described above, we developed a new Box titled, "Description of Programme Objectives, Structural and Curricular Features". Unfortunately, space constraints preclude an adequately comprehensive review of other programmes; however, we believe it is useful for reader to have added the references to ELAM.

### Reviewer 3 Comments and Responses

#### Comment 1

The title does not reflect the objectives, focus and substance of the paper that is concerned with participants' experiences of a specific programme rather than leadership development as such.

#### Response:

We appreciate the reviewer's observation regarding the title, which we revised in response to prior reviewers' feedback. The intention was to better reflect the paper's objective to identify aspects of leadership development programmes perceived as valuable in increasingly complex systems, rather than participant experiences of a specific programme per se. Additional major revisions, made in response to the prior review at BMJQS, were intended to focus the paper in this way as well. We are open to reverting to the original title, *Leadership Development in Complex Health Systems: The Yale and NHS England Experience*. Alternatively, we are very open to any other suggestions and defer to the editor in this regard.

#### Comment 2

Participants' experiences are mainly positive. The interview discussion guide in Box 1 shows no evidence of questions or prompts designed to elicit information on negative experiences or aspects of the program; e.g., What did not work so well for you? Was there anything you would like more of? What aspects do you find challenging to apply in your current practice and context? Should others take the program? Why should they not?

#### Response

We completely agree with the reviewer that prompts to elicit negative experiences or views are essential in qualitative research. We used multiple types of probes to this end, as currently briefly noted (with citations) in the manuscript on page 7: "*Probes were used to generate 'thick descriptions' of their experiences, as well as to elicit both positive and negative views (Creswell & Miller, 2000; Denzin, 1989).*" However, we recognize that clarity on this point is important. We have added the following specific examples of probes both on page 7 of the manuscript and to the discussion guide in Box 2:

Participants were encouraged to share negative views through probes such as: What did not work so well? What was not as worthwhile? What was not as relevant?

We also recognize that there was insufficient exploration of negative findings and accordingly have made two substantive additions to the manuscript. First, as described and included above, we compiled all negative quotes in a new supplemental table titled, “Negative Experiences and Reflections Reported by Participants.” While some quotes have been edited for length, they represent the complete set of all comments with a negative or critical tone in the dataset. We believe this new table is a useful addition and thank the editors and reviewers for this recommendation. The following revision has been made to direct readers to the table on page 8 of the manuscript, as follows:

Some participant feedback included critical or negative comments related to the following general aspects of the programme: networking opportunities, course content, participant selection and cost. All comments are reported in Supplemental Table 2.

And again on page 18 of the manuscript, as follows:

Second, while there are benefits of having the evaluation team familiar with the programme in terms of the quality and rigor of data collection and analysis (see Guba’s Evaluator as Instrument), (Guba & Lincoln, 1981) we recognize the potential risk to objectivity; hence, to ensure transparency we report all negative experience and comments in Supplemental Table 2.

Second, we made the following revision to convey the divergent views of several participants on certain points:

Page 13: First, most, but not all, participants asserted the programme cultivated a robust social network that has provided them with “*common language and common understanding*,” allowing them to increase their impact as they address problems within their systems.

Page 15: Participants reflected that programme learnings ‘*stuck*’ with them (“...*let people have their groups’ is something that possibly every day in my working career since then has stuck with me*”), although ~~one~~ some shared learnings “had worn off a bit”.

Pages 15-16: Of note, one participant described the changes as “incremental” and not as readily perceivable and for another, the relevancy of concepts was not immediately apparent as they wondered “how valuable any of it had been.”

### Comment 3

Participants found certain aspects of the programme to be valuable, but we are told very little about these aspects.

#### Response:

We thank the reviewer for this comment, which is echoed by Reviewer 1 as well. As a major substantive revision, we created a new Box titled, “Description of Programme Objectives, Structural and Curricular Features”. As indicated above, the Box presents comprehensive details on specific structural and design characteristics (with related citations) in order to convey how the programme is operationalized.

### Comment 4

For Theme 1, we are told very little about what the newly-generated scientific evidence is and why it was regarded as so compelling. There is scant detail on the project methodology employed for the field work element. There is also very little detail on the content (concepts, theories, models, framework and so on) that was considered to be especially relevant.

#### Response:

We thank the reviewer for this very helpful comment. We agree that there is limited detail regarding each of these elements, and have made the following revisions, being mindful of space constraints:

**Newly generated scientific evidence:** We have added brief detail about the scientific content and relevant citations in the text as follows on pages 9-10 of the manuscript. We believe the quotations convey why participants found the science compelling ('hot off the press' refers to the fact the article was in pre-publication press at the time hence very timely; the social determinants outcomes and financial data were novel and valuable for one participant's advocacy efforts). We are open to further suggestions if space allows and defer to the editor.

First, the presentation of newly generated scientific evidence relevant to working in complex systems (such as the Leadership Saves Lives study) (Curry et al., 2018) was regarded as particularly compelling since it demonstrated that organizational culture could be improved through intervention: (Curry et al., 2018) *"the research they were giving was hot off the press. That was so exciting."* One participant shared how access to the latest scientific evidence on the role of social determinants (Taylor et al., 2016) of health on health care outcomes and reductions in spending supported her in navigating *"key relationships with our local authorities...the course gave me that evidence, especially trying to relate as much as possible the health and the social component."*

**Field project methodology:** We have added detail about the field project methodology in the new Box "Description of Programme Objectives, Structural and Curricular Features" as follows:

- Participants are supported to:
  - Identify a problem of strategic importance to the organization or system
  - Convene a guiding coalition (Bradley et al., 2018; Kotter, 2001) of diverse stakeholders that can understand and address the problem
  - Facilitate the guiding coalition through the strategic problem-solving process, integrating new perspectives on organizational dynamics
- Faculty feedback provided on deliverables:
  - Problem statement, objective, coalition members, and sponsor sign-off
  - 3-page written summary of project progress and leadership reflections
  - Oral presentation to faculty and peers during module 2

**Concepts, theories, models, frameworks:** The newly created Box titled, "Description of Programme Objectives, Structural and Curricular Features" summarizes core curricular concepts, with citations for readers interested in learning more:

#### *Didactics and table top exercises*

- Working across boundaries (Carlile, 2004; Santos, 2005) and in hierarchy (Jaques, 1990; Ronay, 2012)
- Leadership and followership (Berg, 1990)
- Representational group theory (Alderfer, 1987; Berg, 2005; Curry, 2012)
- Psychological safety (Nembhard & Edmondson, 2006)
- Levels of analysis (Alderfer, 2011; Berg, 2015)

#### *Keynote lectures*

- UK speakers on national strategic priorities and current reforms
- Breaking science from the Yale team's research on leadership, management, culture and organizational performance

- Site visits and roundtables with practitioners/experts on US health and social care

*Strategic problem-solving methodology (Bradley, 2019;Curry, 2013)*

- Defining the problem and objective
- Identifying and prioritizing of root causes
- Generating and pursuing strategic solutions
- Measuring progress and impact

### Comment 5

For Theme 2, we have very little information on the underpinning principles governing network formation how they were facilitated. Nor are the practical tools elaborated.

#### Response

**Network formation.** We agree with the reviewer that there is limited detail regarding network formation. Structural and curricular elements of the programme that support network formation include: a) recruitment of diverse roles, b) a 2-module residential format, c) curricula content focused on working effectively across organizational boundaries, d) strategic problem-solving work sessions and other experiential learning approaches that foster sharing of ideas and perspectives, and e) cultivation of a learning community (opening and closing group sessions, daily reflection). Given space constraints, we have not added this detail to the manuscript, but would be happy to and defer to the editor in this regard. However, to be as responsive as feasible to the reviewer's request, we added 6 quotations about how participants engage in networks after programme completion in Supplemental Table 1 and submitted a revised version as a supplemental file with marked changes for review. Readers are directed to this table for these additional examples on page 13 of the manuscript as follows:

Networks began with intentional facilitation within the programme, have sustained past completion, and expanded across cohorts (See Supplemental Table 1 for quotations describing network formation).

**Practical tools:** We thank the reviewer for this feedback as we expect this practical information might be of interest to readers. Accordingly, we have added brief detail and relevant citations about practical tools in the new Box titled, "Description of Programme Objectives, Structural and Curricular Features", as described and included above. In addition, we have added 5 quotations describing specific tools and techniques participants report using after programme completion in Supplemental Table 1 and submitted a revised version as a supplemental file with marked changes for review. Readers are directed to this table for these additional examples on page 13 of the manuscript as follows:

Second, participants described applying specific tools they learned in the programme to support their work in integrating health and social care across systems (See Supplemental Table 1 for additional examples).

### Comment 6

For Theme 3, very little data is presented to support the theme.

#### Response

We agree with the reviewer and have made two revisions accordingly. We defer to the editor as to whether this is acceptable in terms of the word limit. First, we added the following additional data for Theme 3, on page 15 of the manuscript.

Several described strategies they use to retain learnings and maintain momentum beyond programme completion, ranging from keeping notes and PowerPoints to visual reminders in their workspaces (“Think Yale”) and regular interactions with colleagues:

*I've still got the PowerPoints. With changing jobs, I've kept them, and they come with me. I use them regularly, whether I use them explicitly and say, "Well, I've got this off of Yale" or I use them in terms of having that conversation about, "Well, have you thought of approaching like this? Do you understand what the problem is? Where's your data to justify what the problem is?" (Interview 4 Male, Board Member).*

Second, we added the following example later in this theme's section, on page 16 of the manuscript:

One specific example of a change in perspective was offered by this participant, who shared that her appreciation of the value of diverse teams had grown over time:

*If you've got a team that's all the same, then you need to do something about that team, because there's whole areas that they're not covering. With things like that, my perspective on it has grown. (Interview 14, Female, Senior Manager).*

Third, we have added 2 illustrative quotes to Supplemental Table 1 related to this theme and submitted a revised version as a supplemental file with marked changes for review.

#### **Comment 7**

For the most recent cohort, very little can be inferred from its participants' responses concerning enduring impact and sustainability, given the recency of completion.

#### **Response:**

We agree with the reviewer that little can be inferred from the most recent cohort's responses concerning enduring impact. We checked the data for the 2017 cohort and many reported that they go back to review and use the material, perhaps consistent with their recency of completion. Only 4 participants described learnings as starting to become ingrained, internalized or perceived as more valuable. Most of the reflection on enduring influence is from the 2013 and 2015 cohorts. We added the following clarification to the manuscript on page 15:

Participants (primarily from the 2013 and 2015 cohorts) reflected that programme learnings 'stuck' with them.

#### **Comment 8**

In short, it is hard to discern the contribution that this paper makes to the literature on the design and delivery of leadership development programs. Certain 'core content' areas are delivered over two modules, combined with a field-based project elicited mainly positive feedback from participants at three-time sampling points. However, we know very little about what ideas and tools added what specific value for whom to what extent, why and how. A possible solution to the problems caused by comprising the depth for breadth in this paper might be to devote a paper to each theme and consider what specific knowledge (theories, models, concepts, frameworks), tools and methodologies proved so compelling and valuable for participants' practice, and for their personal and professional development. However, I acknowledge that there may be commercial reasons why the authors do not wish to share this level of detail of curriculum design and delivery, although the references are suggestive. The problem is that without adequate elaboration of key concepts such as 'strategic problem solving' and 'adaptive leadership' it is very hard to appreciate exactly what facets participants found so valuable and why.

#### **Response**

We completely agree with the reviewer that additional detail on each theme would be useful for readers. We hope that the major revisions we have made per the reviewers' excellent comments, as detailed above, have helped to that end. These include: 1) additional detail, clarification and citations in the manuscript text, 2) additional quotations in the manuscript text as space permitted and in supplemental tables, 3) a comprehensive new Box of programme design and curricular features, with citations.

We also appreciate the reviewer's proposed possible solution to develop the themes as separate papers. We carefully considered that solution at length. Because the themes (as well as the components of the programme that support the themes) are so closely interconnected, we are concerned that fragmentation across articles and over time would fail to convey the 'gestalt' of the programme and could leave readers without a holistic understanding.

Finally, we thank the reviewer for noting potential commercial considerations that might preclude sharing details on the curriculum design and delivery. We hope that the new programme features table will be helpful to readers, and our team is fully committed to sharing curricular resources. Related leadership development materials are currently available on the American College of Cardiology website as part of the Leadership Saves Lives study <https://cvquality.acc.org/initiatives/surviving-mi/lsl-toolkit>. One example of this material is the "Essential Tasks of Guiding Coalitions Evidence Brief." This bundle includes a document that briefly describes guiding coalitions, reports main findings and summarizes implications, with a link to the article published in a peer-reviewed scientific journal. Additional resources address three topics: Engaging Diverse Disciplines, Fostering Authentic Participation, and Managing Conflict and Fatigue. Resources include one-pagers for sharing with diverse audiences, descriptive documents, facilitation materials and exercise notes, and PPT decks. Unfortunately, as only members can access the ACC website, we are currently in the process of compiling a resource compendium to make publicly available on our own website at Yale's Global Health Leadership Initiative. Should this manuscript be accepted for publication, we would be glad to add a link to these materials if it is available in time.

## References

- Berg, D. N. (1979). Intergroup relations: an abbreviated update. *Journal of Management Education*, 3, 48-51.
- Bradley, E. H., Brewster, A. L., McNatt, Z., Linnander, E. L., Cherlin, E., Fosburgh, H., . . . Curry, L. A. (2018). How guiding coalitions promote positive culture change in hospitals: a longitudinal mixed methods interventional study. *BMJ Qual Saf*, 27(3), 218-225. doi: 10.1136/bmjqs-2017-006574
- Careau, E., Biba, G., Brander, R., Van Dijk, J. P., Verma, S., Paterson, M., & Tassone, M. (2014). Health leadership education programs, best practices, and impact on learners' knowledge, skills, attitudes, and behaviors and system change: a literature review. *J Healthcare Leadership*, 6, 39-50.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-130. doi: DOI 10.1207/s15430421tip3903\_2
- Curry, L. A., Brault, M. A., Linnander, E. L., McNatt, Z., Brewster, A. L., Cherlin, E., . . . Bradley, E. H. (2018). Influencing organisational culture to improve hospital performance in care of patients with acute myocardial infarction: a mixed-methods intervention study. *BMJ Qual Saf*, 27(3), 207-217. doi: 10.1136/bmjqs-2017-006989
- Dannels, S. A., Yamagata, H., McDade, S. A., Chuang, Y. C., Gleason, K. A., McLaughlin, J. M., . . . Morahan, P. S. (2008). Evaluating a leadership program: a comparative, longitudinal study to assess the impact of the Executive Leadership in Academic Medicine (ELAM) Program for Women. *Acad Med*, 83(5), 488-495. doi: 10.1097/ACM.0b013e31816be551
- Day, D. V. (2000). Leadership development: A review in context. *Leadership Quarterly*, 11(4), 581-613. doi: Doi 10.1016/S1048-9843(00)00061-8
- Denzin, N. K. (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1981). The evaluator as instrument. In E. G. Guba & Y. S. Lincoln (Eds.), *Effective Evaluation*. San Francisco: Jossey-Bass.

Hammersley, M., & Atkinson, P. (1983). *Ethnography: principles in practice*. New York: Tavistock.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2005). *Transferring learning to behavior : using the four levels to improve performance* (1st ed.). San Francisco, Calif.: Berrett-Koehler Publishers.

Kotter, J. P. (2001). What leaders really do. *Harvard Business Review*, 79(11), 85+.

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, 27(7), 941-966. doi: 10.1002/job.413

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3 ed.). Thousand Oaks, Calif.: Sage Publications.

Sudman, S., Bradburn, N. M., & Schwarz, N. *Thinking about answers : the application of cognitive processes to survey methodology* (First edition. ed.).

Taylor, L. A., Tan, A. X., Coyle, C. E., Ndumele, C., Rogan, E., Canavan, M., . . . Bradley, E. H. (2016). Leveraging the Social Determinants of Health: What Works? *PLoS One*, 11(8), e0160217. doi: 10.1371/journal.pone.0160217

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Angus Ramsay University College London, Department of Applied Health Research
<b>REVIEW RETURNED</b>	27-Feb-2020

<b>GENERAL COMMENTS</b>	<p>Many thanks for the opportunity to review the revised version of this paper. Thanks also to the authors, who provided an extremely thorough response to the previous round of reviewer feedback.</p> <p>I feel that our comments have been addressed well: the paper now presents a stronger and more rounded evaluation of the studied leadership development programme. The additional tables of data and feedback offer the reader a clearer sense of how the programme worked and how participants felt about taking part in it.</p> <p>I have only a couple of (minor) points.</p> <ol style="list-style-type: none"> <li>1. Box 2 probably is not necessary in the main manuscript and could be a supplementary file (although the short summary should remain).</li> <li>2. One thought is whether the authors have considered how the programme might develop further in future, especially in terms of sustainability/‘stickability’. For example, is there more that could be done to support participants who struggled to keep hold of the lessons and practices? some sort of long-term follow up? The linkedin groups and so on seem like a really nice way to build ongoing communities of (leadership) practice, and there might be value in thinking more about how these can be used to reduce attrition.</li> </ol> <p>To emphasise, these are both minor points and are just for the authors’ consideration.</p> <p>I look forward to seeing the final version when it is published, and again thank all concerned for the opportunity to support this process.</p>
-------------------------	--

<b>REVIEWER</b>	Roberta E Sonnino MD FACS FAAP Professor of Pediatric Surgery (Retired) Wayne State University School of Medicine Detroit, MI USA Executive, Leadership and Career Coach RES Coaching LLC Locust Hill, Virginia USA
<b>REVIEW RETURNED</b>	25-Feb-2020

<b>GENERAL COMMENTS</b>	<p>The authors have now clarified some elements that were missing in the previous version of the manuscript. The inclusion of the Boxes is extremely helpful in providing curricular details (previously missing) in a well organized manner. Likewise the supplemental tables are an excellent way to present subjective - non quantitative interview findings.</p> <p>The authors also included some references to other courses that address similar leadership training opportunities, which is important: not only do they NOT detract from the value of the author's program, but in fact they provide validation for their methodology.</p> <p>The authors have for the most part modified implications that this program employs methods and principles never used elsewhere, but I would encourage them to be cautious with specific statements such as: "Although leadership development programmes are ubiquitous, NONE have examined participant experiences longitudinally to determine programmatic aspects most useful for leadership in complex systems, and whether or how learnings are retained." There are in fact programs included in the references that have studied these same longitudinal aspects for their own programs over the span of decades... neglecting that other longitudinal studies do exist seems disingenuous. These other reports do not detract from the present study, and some elements of the study are indeed unique.</p> <p>For the rest, this revision addresses the issues I had with the previous manuscript.</p>
-------------------------	--

## VERSION 2 – AUTHOR RESPONSE

### Leadership Development in Complex Health Systems: A Qualitative Study Curry et al

#### Reviewer 1 Comments and Responses

##### Comment 1

I feel that our comments have been addressed well: the paper now presents a stronger and more rounded evaluation of the studied leadership development programme. The additional tables of data and feedback offer the reader a clearer sense of how the programme worked and how participants felt about taking part in it.

##### **Response:**

We sincerely appreciate the reviewers' thorough consideration and suggestions for major refinements of the manuscript. We believe that the paper's contribution has been substantially enhanced as a result.

## Comment 2

Box 2 probably is not necessary in the main manuscript and could be a supplementary file (although the short summary should remain).

### Response:

We have moved Box 2 into a supplementary file as suggested, leaving the short summary within the main manuscript and pointing interested readers to the appendix on page 11 as follows:

All interviews were conducted via phone using a standard discussion guide (see Supplemental Table 1).

## Comment 3

One thought is whether the authors have considered how the programme might develop further in future, especially in terms of sustainability/'stickability'. For example, is there more that could be done to support participants who struggled to keep hold of the lessons and practices? some sort of long-term follow up? The linkedin groups and so on seem like a really nice way to build ongoing communities of (leadership) practice, and there might be value in thinking more about how these can be used to reduce attrition.

### Response

The reviewer raises an important consideration that we continue to explore with our UK partners. We are optimistic about the longevity of the Whats App communities, as we have the past two cohorts quite active and connected in that forum. We have been able to offer "masters classes" once in the past and are eager to develop "booster" (participants' words) programmes in the future. We added the follow brief clause on page 25 of the manuscript:

We hypothesize learnings '*stuck*' and grew over time for several reasons: the intentional creation of a diverse learning community, the integration of leading-edge scholarship into a practical 9-month field project to embed new habits, the curricular balancing of tools and approaches for both strategic thinking and intergroup management,<sup>41</sup> and the resident-style experiences. Future research could explore these hypotheses, as well as the potential value of online communities (such as WhatsApp) in supporting alumnae to retain and apply learnings over time.

## Comment 4

To emphasise, these are both minor points and are just for the authors' consideration.

### Response

Thank you, both are excellent points.

## Reviewer 2 Comments and Responses

### Comment 1

The authors have now clarified some elements that were missing in the previous version of the manuscript. The inclusion of the Boxes is extremely helpful in providing curricular details (previously missing) in a well organized manner. Likewise the supplemental tables are an excellent way to present subjective - non quantitative interview findings.

## Response

We sincerely appreciate reviewers' suggestions to include additional information and clarification in both the manuscript and supplemental files. We are very pleased to hear that these revisions are helpful.

## Comment 2

The authors also included some references to other courses that address similar leadership training opportunities, which is important: not only do they NOT detract from the value of the author's program, but in fact they provide validation for their methodology.

## Response

We thank the reviewer for pointing us to this literature and we agree the inclusion provides validation for the study's methodology.

## Comment 3

The authors have for the most part modified implications that this program employs methods and principles never used elsewhere, but I would encourage them to be cautious with specific statements such as: "Although leadership development programmes are ubiquitous, NONE have examined participant experiences longitudinally to determine programmatic aspects most useful for leadership in complex systems, and whether or how learnings are retained." There are in fact programs included in the references that have studied these same longitudinal aspects for their own programs over the span of decades... neglecting that other longitudinal studies do exist seems disingenuous. These other reports do not detract from the present study, and some elements of the study are indeed unique.

## Response

Then reviewer raises an important caution. We have revised the manuscript as suggested, on page 3 as follows:

Although leadership development programmes are ubiquitous, very few have examined participant experiences longitudinally to determine programmatic aspects most useful for leadership in complex systems, and whether or how learnings are retained.

## Comment 4

For the rest, this revision addresses the issues I had with the previous manuscript.

## Response

We wish to thank the reviewer for donating the time and expertise to such a thorough and constructive review. Our team believes the manuscript has been substantially improved over these revisions and is now much more useful for readers as a result.