SSM - Population Health 7 (2019) 100373

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Article

Developing population health scientists: Findings from an evaluation of the Robert Wood Johnson Foundation Health & Society Scholars Program

Lindsey Realmuto[a], Sheaba Daniel[a], Foram Jasani[a], Linda Weiss[a], Christine Bachrach[b,*]

[a] The New York Academy of Medicine, 1216 5th Ave, New York, NY 10029, USA
[b] Maryland Population Research Center, Morrill Hall, College Park, MD 20742, USA

1. Introduction

As the demand for effective upstream population health interventions grows (Robert Wood Johnson Foundation, 2015; Commission on Social Determinants of Health 2008; House, 2015), so does the need to ensure that scientists are adequately prepared to address the many conceptual, methodological, ethical and logistical challenges inherent in interdisciplinary, multi-level, population health research. Effective training of population health scientists is consequential for virtually everyone involved in population health, whether focused on basic knowledge generation, knowledge translation and transfer, or the development of or advocacy for on-the-ground policies and programs. Even non-scientists who focus on the practice of population health need to rely on interdisciplinary health scientists and the population health research that they produce.

Population health science deviates from traditional science in the sense that it must engage a broad range of disciplines. In addition to mastering the metrics, methods, theories, and body of knowledge in at least one field, trainees also must become broadly knowledgeable about the many other fields – including biology, epidemiology, the social sciences, and more – that must be taken into account to properly design studies and draw conclusions. Trainees also must acquire the orientations and skills needed for interdisciplinary collaboration, such as how to negotiate common understandings and navigate roles and responsibilities in team science. Ideally, trainees must learn how to design research agendas, conduct studies, and communicate research in ways that facilitate the use of knowledge in policy and practice settings (Bachrach, Robert, Green, Shostak, Thomas, & 2015).

However, the training of population health scientists has received little systematic attention. Although guidelines for training in public health are well-established (Council on Education for Public Health, 2016), the goals of population health go beyond the traditional public health model (Valles, 2018). In their seminal textbook, Keyes and Galea (2016:94) call for "A critical self-reflection on how we train and socialize scholars" and note that this "will be paramount to enacting an approach that focuses attention on curve-shifting in population health."

In 2015, the National Academy of Medicine's Roundtable on Population Health Improvement commissioned a report based on a two-day meeting of scientists involved in population health training. The report suggested three essential elements for population health science training: immersion of trainees in an interdisciplinary environment; mentoring; and experience as part of an interdisciplinary research team (Bachrach et al., 2015). Such expert opinion provides a starting point for discussion, but a field accustomed to relying on empirical evidence should go further. How do we know what works? This question may be especially relevant to a field that is attempting to work within new paradigms.

To our knowledge, no other published literature addresses whether and how training programs can meet the broader needs of population health science. Indeed, there is little published research on the outcomes of any postgraduate public health program, and that which exists is largely characterized by weak designs. A 2013 literature review found eight studies conducted between 1995 and 2012 that evaluated career outcomes for trainees completing full-time public health programs. Only one of these included a comparison group and only three included statistical tests of significance. The one study with a comparison group was limited to 18 trainees and 10 unselected applicants (Faupel-Badger, Nelson, Marcus, Kudura, & Nghiem, 2013). More recent studies include several evaluating the National Cancer Institute's Cancer Prevention Fellowship Program (Faupel-Badger, Raue, Nelson, & Tsakraklides, 2015; Faupel-Badger, Nelson, & Izmirlian, 2017), including one study with a strong comparison group and methods (Faupel-Badger et al., 2015). A few recent studies evaluate the outcomes of other public health training programs but lack comparison groups (Sobelson, Young, Wigington, & Duncan, 2017; Baldwin et al., 2017; Mancuso et al., 2017).

This article is intended to begin a conversation about training in population health science. It describes the Robert Wood Johnson Foundation (RWJF) Health and Society Scholars (HSS) program, a post-doctoral program that operated over a period of 12 years; it compares the post-program research trajectories and scientific impact of trainees with that of a closely matched comparison group; and it highlights the need to study programs to learn what works, even in the (sensible) absence of random assignment of training slots. By providing

* Correspondence to: Christine Bachrach, 6253 Cardinal Ln, Columbia, MD 21044, USA.
E-mail addresses: chrisbachrach@gmail.com, bachracc@umd.edu (C. Bachrach).

https://doi.org/10.1016/j.ssmph.2019.100373
Received 26 October 2018; Received in revised form 29 January 2019; Accepted 2 February 2019

quantitative data on the outcomes of one training program, it invites attention to the outcomes of others and an ongoing exchange about best practices in the training of population health scientists.

1.1. Design of RWJF Health & Society Scholars

The HSS program was launched in the face of multiple perceived challenges: the early stage of the field, disciplinary silos, the need for integration of disciplines that embraced different methodological approaches, and uncertain job markets for population health scholars (Harper, 2016). The primary goal was to generate scientific expertise and productivity, with capacity for leadership and the translation of research to policy and practice as additional goals.

The program was designed collaboratively by RWJF staff and the directors of the six sites[1] selected to participate in the program. At each site, and over a period of two years, trainees from diverse disciplinary backgrounds interacted intensively in seminars, working groups, and collaborative research projects. They engaged with faculty from multiple disciplines including medicine, public health, the social sciences, and in some cases, business, education, and law. They received mentoring from faculty members about interdisciplinary skills and career development. Trainees, faculty, and alumni from all sites gathered at an annual conference to share research, engage in interdisciplinary exchanges, and hone presentation skills. Supported by a National Advisory Committee and RWJF staff, the HSS National Program Office monitored activities at the six sites, held meetings of site directors at annual meetings, and conducted site visits on an annual or biennial basis. A qualitative assessment of program reports found that all sites were implementing the program consistently and effectively but that, especially during the early years of the program, training in leadership and translation was less consistent (Harper, 2016; Hiatt, 2018).

Between 2003 and 2016, HSS trained 193 early-career scientists. The program is widely viewed as successful. It attracted a large number of talented applicants from a diversity of disciplines. Alumni, faculty, and advisors have all reported that the program had a distinct impact on their scientific approaches, their careers, and their institutions (Harper, 2016). However, such reports provide insubstantial evidence of impact with respect to desired program outcomes. This evaluation was undertaken to produce such evidence.

2. Materials and methods

2.1. Sample

The sample of program participants included all 90 HSS alumni from cohorts 2–6 (admitted during 2004–2008). The evaluation focused on the program's early to middle cohorts to allow for assessment of outcomes for a sufficient duration of time. Cohort 1 was excluded because data were incomplete and/or non-comparable to the later cohorts. The comparison group was a purposively selected subsample of finalists for the program in the same years.

HSS had a rigorous application and selection process to ensure that the individuals with the best qualifications and fit for the program were enrolled. This non-random process poses significant challenges for measuring program impact. If programs admit only the best and the brightest, then it is impossible to tell whether measured outcomes can be attributed to the program or the pre-existing applicant characteristics that prompted admittance to the program. This paper uses two strategies that have been shown to mitigate this problem (Pion & Cordray, 2008; Faupel-Badger et al., 2015): first, the selection of a comparison group that is as similar as possible to program participants on characteristics relevant to the selection process and, second, the use of statistical methods to control for any remaining differences between trainee and comparison groups. These strategies were possible because HSS collected and retained a wealth of data on each applicant, including a quantitative rating summarizing their qualifications and fit for the program. The selection problem cannot be fully mitigated, however, because it is impossible to verify that all factors relevant to both program selection and observed outcomes have been measured and controlled. Thus, findings must be viewed as suggestive and not conclusive.

Identifying a comparison group from applicants to the program establishes, at the outset, some level of similarity in career goals and interests (Faupel-Badger, Nelson & Izmirlian, 2015; Pion & Cordray, 2008). Similarity in qualifications for the program can be further optimized by choosing comparison cases from among the applicants who came closest to being chosen themselves, in this case, program finalists.

Fig. 1 shows the selection process. The program received a total of 859 applicants to the program for cohorts 2–6. Each year, an initial screening by program staff and outside consultants reduced the pool of applicants by nearly one-half, retaining as semi-finalists those considered to have the best qualifications and fit for the program on the basis of curriculum vitae (CVs), academic records, references, personal essays and writing samples. Potential for leadership, initiative, creativity and academic success were considered in these decisions, as well as the need for disciplinary diversity in the program. In total, 442 cohort 2–6 applicants were retained as semi-finalists.

Qualifications of semi-finalists were reviewed by the HSS National Advisory Committee (NAC) in order to develop a finalist pool. Using the same criteria and materials as in the initial screening, four NAC members independently evaluated each of the semi-finalists, assigning scores (range 1–10) to reflect the applicant's qualifications and fit for the program. Average NAC score was a primary determinant of finalist selection, supplemented by discussion of candidates that fell at the margin or received highly divergent scores at an in-person meeting of the NAC. A total of 253 individuals were chosen as finalists in cohorts 2–6.

Onsite interviews were the final step in the selection process and most finalists were interviewed by at least one program site. However, 33 voluntarily withdrew from consideration to pursue other opportunities, in most cases prior to having interviews. Fourteen of these decided to interview for the RWJF Scholars in Health Policy Research program and were not allowed to interview for both programs. Nineteen withdrew because they had accepted other offers. In addition, nine interviewed candidates declined offers from the program, bringing the total number of finalists who self-selected out of the program to 42 (17% of the finalist pool).

The comparison group was purposively drawn from the pool of all finalists who did not participate in the program (N = 163), including those who withdrew.[2] It was designed to be as similar as possible to the group placed in the program on characteristics associated with program participation. To identify relevant characteristics, we examined whether, among all finalists, NAC score, gender, discipline, type of doctoral degree, receipt of other postdoctoral training, and receiving an undergraduate degree from a top 10 liberal arts/research institution predicted participation in the program. Although none of these had statistically significant effects on selection, NAC score was most strongly related ($p = .11$).[3]

Given these results, we selected a subsample of 117 of the non-enrolled finalists who most closely matched alumni on year of application

[1] Program sites were located at Columbia University, Harvard University, the University of California - San Francisco/Berkeley, the University of Michigan - Ann Arbor, The University of Pennsylvania, and the University of Wisconsin-Madison Sites differed in their faculty, interdisciplinary strengths, and major emphases, but the key features of the program described here were common across all sites. While an examination of between-site differences could be of interest, it is beyond the scope of this paper.

[2] These individuals were retained in the pool for the comparison sample because they tended to have high qualifications for the program and clearly had interests in population health.

[3] Results available upon request.

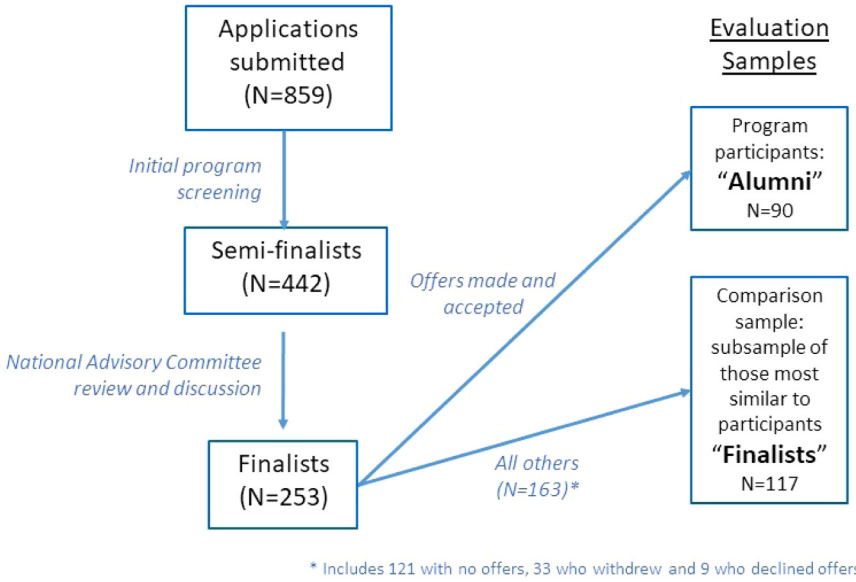

Fig. 1. Selection Process for Health & Society Scholars Cohorts 2–6 and Construction of Evaluation Samples.

(cohort), NAC score, gender, and field of study. The number of finalists selected was higher than the number of alumni to allow for higher nonresponse among program finalists. Two finalists were excluded from the final sample: one refused and one could not be located.

2.2. Period of observation

The goal of the evaluation was to understand whether HSS influenced alumni careers and contributions after leaving the program, when publications and other accomplishments would be more clearly tied to individuals' own agendas than to specific program-based activities. The sampled alumni left the program between 2006 and 2010; measures of professional outcomes were assessed for the period 2011–2015. This allowed five full years of observation for all individuals,[4] but meant that cohort outcomes were measured at differing intervals post-program. Cohort 2 outcomes were measured 5–10 years after program exit; those for cohort 6, 1–5 years after. These variations in post-program observation periods were addressed in two ways: by selecting a sample of finalists that closely matched program alumni in the distribution across cohorts; and by controlling for cohort and years since doctoral degree in the analyses. As these two factors were closely correlated, only the latter was retained in multivariate analyses.

2.3. Sources of data, data collection and response rates

Previous training program evaluations have relied on data from a variety of sources, including archival records, surveys, CVs, on-line searches, and bibliographic databases, with most studies relying on a single source of data (Faupel-Badger et al., 2013; Faupel-Badger, Nelson & Izmirlian, 2015). Completeness and utility of data vary by source. Archival records tend to be complete but cannot measure post-program outcomes. Surveys offer the opportunity to measure trainee perceptions and other self-reported information but may suffer from social desirability bias and, especially when conducted after subjects have left the program, low response rates; response rates for comparison subjects may be especially low (Mancuso et al., 2017; Faupel-Badger, Nelson & Izmirlian, 2015). CVs must be obtained directly from subjects or through on-line searches; while these reliably include positions held and certain types of publications (e.g., journal articles, books), other desired information is not consistently reported.[5] On-line searches can identify positions held in many cases; however access to other information varies widely. Bibliographic databases provide an independent record of published articles and other indicators of scholarly success but all databases are incomplete in some way (De Groote & Raszewski, 2012; Mongeon & Paul-Hus, 2016).

Because no one source is perfect, this study gathered data from multiple sources:

(1) Archival records from the HSS application process were available for all alumni and finalists.
(2) CVs (current within two years) were obtained for 92% of alumni and 65% of finalists through direct requests and online searches. For cases missing CVs, information was obtained from online searches, with information from the original application used to verify identity. These searches allowed us to identify current position for all subjects and to characterize past positions for all alumni and 91% of finalists.
(3) Bibliometric data (i.e., statistics on published articles) were derived from a Web of Science (WoS) search for the years 2011-15. The search included multiple names (e.g., original and married names) for individuals. Individual identities were verified by comparing WoS data with publications and affiliations listed on recent CVs, websites, and application materials. A total of 2,544 journal articles written by alumni and finalists from 2011 to 2015 were identified, with publications found for 97% of alumni and 86% of finalists. The difference in these percentages was consistent with CV information on publications for the two groups.
(4) A brief online survey was also conducted to supplement information from other sources. Participants were contacted by email and phone between April and October, 2016, to encourage participation. The survey was administered online with a link sent to participants via email. Survey responses were obtained from 88% of alumni and 63% of finalists.

Information was obtained from at least one source for all of the alumni and finalists retained in the sample. The study protocol was approved by the Institutional Review Board at [blinded for review].

[4] Constraints on the timing of funding for the evaluation prevented a longer period of observation.

[5] Based on analysis of the 158 CVs collected as part of this study.

2.4. Measures

This report focuses on three sets of measures, selected for consistency with HSS program objectives and data quality. The first set included "baseline" data: year applied to the program, the highest degree(s) and year degree was received, discipline, gender, and NAC rating. These data were drawn from application materials and HSS records.

A second set, professional outcomes, included professional positions, leadership, policy involvement, and professional output. We identified current position and positions held in the past five years ("career trajectory"), differentiating employment in academia, outside academia, and in mixed settings, as well as position within academia and type of nonacademic work (research vs. other) when relevant. Two types of leadership positions (held within the past five years) were coded from CV data: professional (those connected to paid work, such as directing an office or center) and extramural (e.g., serving on a board, holding a position in a professional organization). The survey ascertained self-reported measures of leadership experience related to population health and engagement in population health policy or practice (e.g., disseminating the results of research to a policy maker).

Measures of professional output included the number of external grants since 2011 (from CVs) and two measures drawn from WoS: the number of journal articles published during 2011–2015, and the Hirsch h-index score (which indicates how frequently publications had been cited).

A third set of variables measured engagement in population health research. The first of these variables was based on a survey question regarding engagement in population health research in the last five years. The second was derived by selecting and coding, for each alumnus or finalist, a sample of up to five journal articles published during 2011–2015 and listed in Web of Science. For individuals with five or fewer articles, all articles were included; for those with more than five, a sample was drawn using a random number generator. Because alumni were more likely than finalists to have five or more articles in WoS, the average number of articles coded was higher for this group: 4.4 compared to 3.8 for finalists.

Articles were considered "population health research" if they addressed levels of or differences in health in populations and/or addressed the linkages between social/environmental factors and health.[6] Coding was based on the publication's title, supplemented by reviews of publication abstracts. At least two members of the research team coded each publication, with disagreements adjudicated by the full group. The coding was blinded to reduce potential bias. Two summary measures were created for each individual in the sample based on this coding: (1) any articles coded for having population health content and (2) the number of articles so coded.

2.5. Data analysis

Alumni and finalists were compared using bivariate and multivariate analyses. In bivariate analyses, the significance of alumni-finalist differences was assessed using chi-square and t-tests. Multivariate analyses were conducted where there were significant differences between alumni and finalists in the bivariate analysis, in order to confirm that differences were independent of measured pre-selection differences between groups. Multivariate models included logistic regression (binary dependent variables) and Poisson and negative binomial regression (count variables). Multivariate models controlled for NAC score, discipline (grouped into health sciences, social sciences and other), and years since doctoral degree. Models were also run with controls for current position (i.e., academic or not) to examine the role of position in mediating any differences associated with program participation.

[6] Detailed coding instructions are available upon request.

3. Results

Table 1 describes basic characteristics of the sample, including gender, cohort, highest degree(s), and discipline. None of these characteristics differed significantly between alumni and finalist groups. Both groups were two-thirds female and mainly received PhDs as their highest degree. There was a slight tendency for alumni to have received their doctorates more recently. Similar proportions of finalists and alumni had degrees from the social sciences (46–47%) and the health sciences (public health and medicine, 42%); 11–12% had degrees from a variety of other disciplines, including the biological sciences, business, engineering, and others. Alumni and finalists did not differ in their qualifications and fit for the program as indicated by the mean NAC scores assigned during the application process (8.0 vs 7.9 respectively).

Table 2 describes and compares professional outcomes. None of the differences in current position were statistically significant. However, when career trajectories over the past five years were examined, alumni were more likely to have held positions in both academic and non-academic settings ("Mixed") and finalists were more likely to have held positions outside of academia only. About two-thirds of those holding non-academic positions were conducting research.

Differences between alumni and finalists with respect to leadership were nonsignificant, although alumni had a greater tendency than finalists to report leadership on the survey (60% vs. 49%). In response to a survey question on policy involvement, an equal proportion of alumni and finalists reported that they had been engaged in population health practice or policy within the past five years.

Between 2011 and 2015, alumni had, on average, a greater number of journal articles identified in WoS compared to finalists (15.2 versus 10.3, $p = .014$). Alumni also had a significantly higher Hirsch h-index average: 5.74 compared to 3.96 for finalists ($p = .001$), indicating greater publication impact. In multivariate analyses (not shown), alumni-finalist differences in both of these measures remained significant, and this result held even when current position (academic/other) was included in the model. Differences between alumni and finalists in the number of external grants received were not significant.

Results of the publication coding indicate that alumni were more involved in population health research than finalists. Published articles by alumni were more likely to be coded as population health research than were those by finalists (64.1% vs 51.0%, $p < .001$). On the survey, 91% of alumni versus 82% of finalists ($p = .041$) reported engagement in population health research (see Table 3). Bibliometric data show similar differences: 87.8% of alumni and 69.5% of finalists published at least one article coded as population health research during 2011–2015 ($p = .002$, see Table 4). The average number of population health articles was higher for alumni (2.82 vs 1.93; $p < .001$). In multivariate analyses, both differences remained highly significant.

This result does not seem to be a function of the higher number of articles coded for alumni. If we restrict our comparison to alumni and finalists with 5 or more WoS articles, we find that, in the random sample of 5 articles coded for each, alumni published 4.5 pop health articles to the finalists' 2.5. Further, as shown in Table 3, the *proportion* of coded articles that addressed population health was also higher for scholars.

4. Discussion

The evaluation results show significant differences between alumni and finalists on measures relevant to the central goal of HSS – to produce productive scientists contributing to population health research. In the absence of random assignment to the program, it is impossible to determine conclusively whether the differences result from selection or program effects. While the steps taken in this study to address potential threats to inference bolster confidence in the validity of the results, limitations must be taken into account.

Table 1
Participant characteristics*.

	Alumni N (%)	Finalists N (%)	Total N (%)	P value
Gender				
Female	60 (66.7)	74 (64.3)	134 (65.4)	.729[e]
Male	30 (33.3)	41 (35.6)	71 (34.6)	
Cohort				
2	15 (16.7)	23 (20.0)	38 (18.5)	.906[e]
3	21 (23.3)	22 (19.1)	43 (21.0)	
4	19 (21.1)	22 (19.1)	41 (20.0)	
5	17 (18.9)	25 (21.7)	42 (20.5)	
6	18 (20.0)	23 (20.0)	41 (20.0)	
Average Year Completed Doctorate	2005.0	2004.1	2004.5	.076[f]
Highest Degree				
Dr.P.H.	2 (2.2)	0 (0.0)	2 (1.0)	.148[e]
D.Sc.	11 (12.2)	6 (5.2)	17 (8.3)	
M.D.	5 (5.6)	8 (7.0)	13 (6.3)	
M.D./Ph.D.	0 (0.0)	1 (0.9)	1 (0.5)	
Ph.D.	72 (80.0)	100 (87.0)	172 (83.9)	
Discipline				
Health Sciences[c]	38 (42.2)	48 (41.7)	86 (42.0)	.970[e]
Social Sciences**	41 (45.6)	54 (47.0)	95 (46.3)	
Other[d]	11 (12.2)	13 (11.3)	24 (11.7)	
Total	**90**	**115**	**205**	

* Measures drawn from initial application to program and reviewer scoring of applicants.

** Social science degrees included: anthropology/medical anthropology, demography, economics, geography, history, political science, psychology, public policy, social work, and sociology.

[c] Health Science disciplines included: epidemiology, environmental health, health behavior, health policy, medicine, and public health.

[d] 'Other' disciplines included: biological sciences, business, communications, engineering, neurosciences, and urban planning.

[e] Chi-square statistic.

[f] T-test statistic.

Table 2
Professional outcomes, 2011–2015*.

		Alumni N (%)	Finalists N (%)	P value[e]
Professional Position				
Current position (N = 205)	Academic Only	57 (63.3)	72 (62.6)	.093
	Mixed	19 (21.1)	14 (12.2)	
	Non-Academic Only	14 (15.6)	29 (25.2)	
Career Trajectory (over past 5 years)	Academic	50 (55.6)	66 (57.4)	.033[d]
	Mixed	30 (33.3)	24 (20.9)	
	Non-Academic	10 (11.1)	19 (16.5)	
Position within academia (N = 162)**	Tenured	37 (48.7)	43 (50)	.319
	Tenure Track, not tenured	28 (36.8)	24 (27.9)	
	Other	11 (14.5)	19 (22.1)	
Role outside of academia (N = 43)***	Nonacademic Research	10 (71.4)	19 (65.5)	.698
	Other	4 (28.6)	10 (34.5)	
Leadership				
Professional		30 (35.7)	30 (40.0)	.367
Extramural		25 (29.8)	22 (29.3)	.634
Leadership Relevant to Population Health Research in Last Five Years		47 (59.5)	36 (49.3)	.058
Policy Involvement				
Engagement in population health practice or policy in the past five years		54 (69.2)	50 (68.5)	.922
Professional Output				
		Mean	**Mean**	**P value[f]**
# Journal articles published (2011–2015)		15.18	10.3	.014[d]
Hirsch h-index score		5.74	3.96	.001[d]
Disciplines cited per article		9.00	7.84	.000[d]
# of External grants awarded		5.14	4.48	.350

* Data sources for measures: (1) Professional position: CVs; current websites when CVs unavailable; (2) Leadership: CVs; survey; (3) Policy Involvement: Survey; (4) Professional Output: Web of Science bibliographic search.

** Limited to participants with a current academic position.

*** Limited to participants working outside of academia.

[d] p value < .05.

[e] Chi-square statistic.

[f] T-test statistic.

The most obvious threat to inference is selection bias. We used a two-pronged strategy to minimize selection effects. First, the comparison group included only individuals who were finalists for the program, interviewed for a slot, and whose ratings on qualifications and fit for the program were similar to those of program alumni. Second, we used multivariate methods to control for factors with non-negligible associations with enrollment in the program. It is impossible to know whether other, unmeasured, factors could have influenced both selection for the program and the outcomes. Although qualitative accounts suggest that selection by the sites was guided primarily by a desire to create groups of scholars with diverse disciplinary backgrounds but complementary interests, in-person interviews could have revealed traits not captured in the NAC scoring that influenced both admission and later success. Thus, the possibility of selection bias in our results cannot be eliminated.

A major strength of the evaluation was the incorporation of a wide variety of measures derived from several sources. This approach facilitated examination of multiple outcomes of the program while ameliorating challenges resulting from missing data. As in other studies of this type, finalists were less likely than alumni to respond to the survey or provide a CV (Faupel-Badger, Nelson & Izmirlian, 2015). The findings that rely on these data (e.g., leadership, policy involvement, grants, and self-reported involvement in population health research) may well be affected by response bias. Only one of the statistically significant results pertains to these outcomes (self-reported involvement in population health research), and this result was confirmed by an analysis based on bibliographic data.

The main differences that emerge between alumni and finalists relate to publication records, and these are drawn from bibliographic data (WoS). Like other bibliographic databases, WoS does not capture the full universe of published articles and does not include scholarly output in books, book chapters, or technical reports. Analyses of WoS coverage show it to be slightly biased against inclusion of social science publications (Mongeon & Paul-Hus, 2016). However, there is no reason to believe that WoS differentially captured publications from alumni and finalists, given their virtually identical composition by discipline. A similar 3:2 ratio was found between the numbers of publications cited in CVs by alumni and finalists, further buttressing the validity of the WoS results.

A weakness of the study, unavoidable given the short duration of the program, was the relatively small sample size. In many cases, the results approached, but did not reach, statistical significance. Coding only a sample of publications for population health content may have decreased reliability, but the likelihood of bias was minimized by randomly selecting the publications coded.

It is possible that the differences between alumni and finalists are attributable to differences in career trajectories. Finalists were more likely than alumni to have held only non-academic jobs over the study period. The measures on which alumni and finalists differed significantly are related to their publications, and publishing is most strongly valued in academic settings. In multivariate analyses, we examined whether holding an academic position accounted for the relationships we found, but found no evidence of this.

Because the samples of alumni and finalists were, by design, equally promising researchers at the time of application to the program, they

Table 3
Engagement in Population Health Research.

	Alumni N (%)	Finalists N (%)	P value[a]
Engagement in population health research (survey)	72 (91.1)	60 (82.2)	.041[b]
Number of published articles considered population health research	254 (64.1)	223 (51.0)	< .001[b]

[a] Chi-square statistic calculated.
[b] p value < .05.

Table 4
Summary of individual-level indicators of engagement in population health research – Bivariate and Multivariate Analyses.

	Bivariate Analyses			Multivariate Analyses			
	Alumni	Finalist	P value	Odds ratio*	95% CI	B**	95% CI
Any article coded for population health content	79 (87.8%)	80 (69.5%)	0.002[c,d]	2.86[c]	1.33–6.15		
Number of articles coded for population health content (mean)	2.82	1.93	< .001[c,e]			1.368[c]	1.14–1.64

* Logistic regression multivariate model; controlled for discipline, years since doctoral graduation and NAC score.
** Poisson Regression multivariate model; controlled for discipline, years since doctoral graduation and NAC score.
[c] p-value < .05.
[d] Chi-square statistic.
[e] T-test statistic.

could be expected to be similar on standard measures of productivity and professional advancement. Indeed, program alumni and finalists did not differ on many of the outcomes we examined, including tenure, external funding, and other aspects of current position. Alumni did, however, outperform finalists in both numbers of publications and publication impact scores.

One possible interpretation of these differences is that they are a function of the protected time afforded by having a postdoctoral fellowship. However, a comparison of the number of WoS publications for the 53 finalists who had completed postdoctoral programs other than HSS with those for the remaining finalists and program alumni suggests otherwise. For example, finalists with and without other postdocs had virtually identical numbers of publications (10.1 and 10.5, respectively) and the differential with program alumni (15.2) was unchanged. Analysis of publications listed on CVs produced similar results.

The Robert Wood Johnson Foundation invested substantial resources in HSS. Program participants were paid salaries above the norm for post-doctoral fellows; mentors were paid to mentor; funds were available for research and travel to meetings. Even if one accepts the findings of this study as evidence of program success, it remains impossible to identify which specific features of the program mattered most: was it these resource-intensive features or other elements of the program design that could be replicated in other, more modestly funded, postdoctoral programs?

Interviews conducted by RWJF with alumni, faculty, and observers of the program (Harper, 2016) produced qualitative information on the value of specific program features. Some of the features identified – seminars, working groups, mentoring, and supplemental funding for research and travel – are common among other postdoctoral programs. However, funding for mentoring is not, and most programs do not hold annual conferences to bring trainees together; these features were often cited as critical to the program's success. While the high salaries paid to participants were an important draw for talented applicants, it is unclear whether these could have affected program impact.

One area in which the program apparently fell short of its goals was in increasing leadership activity and policy involvement among its alumni. This may reflect problems of data quality or, more likely, the restriction of the study to the earliest cohorts, when these goals were less often, and less consistently, addressed. Only one site consistently emphasized translational issues, and few provided leadership training (Harper, 2016). Towards the end of the program, efforts were made to improve training in these areas, but those engaged at that time (i.e., the later cohorts) were not included in this study.

This paper did not address whether alumni and finalists differ in their involvement in interdisciplinary research; a separate analysis in progress examines this question in depth using bibliometric approaches. Preliminary evidence suggests that alumni are significantly more likely than finalists to produce work that draws from multiple disciplines, and substantially more likely to produce such work in studies that focus on population health (Bachrach et al., 2017).

5. Conclusions

Tackling challenges in population health requires researchers and practitioners who are able to address complex, multidisciplinary issues. It is likely that the most important advances in this field will come from integrating high-level contributions from disciplines that address health determinants at multiple levels, suggesting the need to engage scientists with advanced training in the relevant disciplines. These considerations underscore the need for interdisciplinary population health training at the post-doctoral level, when disciplinary skills have been fully acquired. The HSS program was created to meet that need. This evaluation provides information supporting its success, although it cannot provide conclusive evidence.

In most areas of public health investment, outcomes evaluation is a given. Yet, there is a dearth of evidence on the outcomes of training programs and none addressing viable models for meeting the specific goals of population health science. No doubt, the lack of random assignment in such programs is a deterrent to outcomes evaluation. We argue that studies can nevertheless provide relevant information if comparison groups are rigorously constructed. This, in turn, depends on early investments in evaluation by program developers, including forethought about the key questions to answer, measurement of factors influencing the selection process, and documentation of applicants' pre-program skills and accomplishments. The resulting studies can help to build knowledge about what training models effectively advance population health science, and how and why they are successful.

Acknowledgements

We thank the Robert Wood Johnson Foundation for their support of this research (Grant # 72968); Tongtan Chantarat, Gerard Lebeda, and Jeffrey Price for their many contributions in the beginning stages of the study; Kristin Harper, Dave Kindig, Pamela Russo, and Jonathan Samet for advice on study design and data analysis; and two anonymous reviewers for helpful comments.

Declaration of interest statement

Funding for this work was provided by the Robert Wood Johnson Foundation (RWJF) at the closing of Health & Society Scholars (HSS) in 2016. Author Christine Bachrach was previously Co-Director of the National Program Office for HSS. Pamela Russo of RWJF served on the project's Advisory Group and participated in encouraging responses from study participants. There was no other involvement in study design, implementation, analysis or reporting on the part of RWJF staff. The authors have no financial conflicts of interest to disclose at this time.

References

Bachrach, C. A., Robert, S., Green, T., Shostak, S., & Thomas, Y. (2015). *Training in Interdisciplinary Health Science: Current Successes and Future Needs*. Washington DC: Institute of Medicine Roundtable on Population Health Improvement. (Accessed February 6, 2019) http://iom.nationalacademies.org/~/media/Files/Activity%20Files/PublicHealth/PopulationHealthImprovementRT/Commissioned%20Papers/Training%20Population%20Health%20Science%20final.pdf.

Bachrach C.A., Moody J., Sheble L., et al. (2017). Effects of an Interdisciplinary Postdoctoral Program on Interdisciplinary Science. In: Science of Team Science (SciTS) 2017 Conference. Clearwater Beach, FL. ⟨http://www.scienceofteamscience.org/2017-scits-conference⟩.

Baldwin, J. A., Williamson, H. J., Eaves, E. R., et al. (2017). Broadening measures of success: Results of a behavioral health translational research training program. *Implementation Science, **12***, 92. https://doi.org/10.1186/s13012-017-0621-9.

Commission on Social Determinants of Health (2008). Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health. Final Report of the Commission on Social Determinants of Health. Geneva, Switzerland.

Council on Education for Public Health (2016). Accreditation Criteria Schools of Public Health & Public Health Programs: Amended October 2016. ⟨https://storage.googleapis.com/media.ceph.org/wp_assets/2016.Criteria.pdf⟩. (Accessed 14 December 2018).

De Groote, S. L., & Raszewski, R. (2012). Coverage of Google Scholar, Scopus, and Web of Science: A case study of the h-index in nursing. *Nursing Outlook, 60*(6), 391–400. https://doi.org/10.1016/j.outlook.2012.04.007.

Faupel-Badger, J., Nelson, D. E., Marcus, S., Kudura, A., & Nghiem, E. (2013). Evaluating Postgraduate Public Health and Biomedical Training Program Outcomes: lost opportunities and renewed interest. *Journal of Cancer Education, 28*(1), 18–26.

Faupel-Badger, J. M., Raue, K., Nelson, D. E., & Tsakraklides, S. (2015). Alumni perspectives on career preparation during a postdoctoral training program: A qualitative study. *CBE–Life Sciences Education, 14*(1), 1–8. https://doi.org/10.1187/cbe.14-06-0102.

Faupel-Badger, J. M., Nelson, D. E., Izmirlian, G., et al. (2015). Independent association of postdoctoral training with subsequent careers in cancer prevention. *PLoS One, 10*(12), 1–11. https://doi.org/10.1371/journal.pone.0144880.

Faupel-Badger, J. M., Nelson, D. E., & Izmirlian, G. (2017). Career satisfaction and perceived salary competitiveness among individuals who completed postdoctoral research training in cancer prevention. *PLoS One, 12*(1), e0169859. https://doi.org/10.1371/journal.pone.0169859.

Harper, K. (2016). Building the Science of Population Health: The Robert Wood Johnson Foundation Health & Society Scholars Program 2001–2016 and Beyond. ⟨http://www.healthandsocietyscholars.org/media/file/LessonsLearned-Report12-17-16FULL.PDF⟩. (Accessed 17 October 2018).

Hiatt, R.A. (ed). (2018). Population Health: The Translation of Research to Policy. New York: Milbank Memorial Fund ⟨https://www.milbank.org/wp-content/uploads/2018/09/MMF-HHS-The-Case-Studies-all.pdf⟩. (Accessed 27 December 2018).

House, J.S. (2015). Beyond Obamacare: Life, Death, and Social Policy. New York, Russell Sage.

Keyes, K. M., & Galea, S. (2016). Setting the agenda for a new discipline: Population health science. *American Journal of Public Health, 106*(4), 633–634. https://doi.org/10.2105/AJPH.2016.303101.

Mancuso, J. D., Garges, E. C., Hickey, P. W., et al. (2017). Outcomes from US military-supported overseas training rotations in tropical medicine and global health, 2006–2015. *Military Medicine, 182*(9), e1796–e1801. https://doi.org/10.7205/MILMED-D-16-00408.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics, 106*(1), 213–228.

Pion, G. M., & Cordray, D. S. (2008). The Burroughs Wellcome Career Award in the biomedical sciences: Challenges to and prospects for estimating the causal effects of career development programs. *Evaluation & The Health Professions, 31*(4), 335–369. https://doi.org/10.1177/0163278708324434.

Robert Wood Johnson Foundation (2015). From Vision to Action: A Framework and Measures to Mobilize a Culture of Health. Princeton, NJ: Robert Wood Johnson Foundation.

Sobelson, R. K., Young, A. C., Wigington, C. J., & Duncan, H. (2017). Early evaluation findings from a federally funded training program: The Public Health Associate Program. *Journal of Public Health Management and Practice, 23*(5), 439–446. https://doi.org/10.1097/PHH.0000000000000546.

Valles, S. A. (2018). *Philosophy of population health.* New York: Routledge.