HHS Public Access

Author manuscript

Eval Program Plann. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:

Eval Program Plann. 2020 June ; 80: 101812. doi:10.1016/j.evalprogplan.2020.101812.

An exploratory review of the literature evaluating nonclinical fellowship programs

Margaret Paek*, **Connor Radkey**, **Sally Honeycutt**, **M. Kathleen Glynn**

Centers for Disease Control and Prevention, Office of Public Health Scientific Services, Center for Surveillance, Epidemiology, and Laboratory Services, Division of Scientific Education and Professional Development, 1600 Clifton Road NE, Atlanta, GA 30333, United States

Abstract

Fellowship programs offer career development opportunities, provide experiential training, and can be used to recruit personnel to address specific challenges facing the public health workforce. Given the potential influence fellowships have on the future public health workforce, it is important to understand and articulate the results of such programs and to identify areas of improvement to meet current workforce needs. The purpose of this literature review was to identify common practices used to evaluate nonclinical fellowship programs. After a search of the internet and selected databases, we screened titles and abstracts using predetermined selection criteria. We then conducted a detailed review of selected papers to extract information about program characteristics (program description, sector, and program length) and evaluation characteristics (primary evaluation type, framework for evaluation, data collection methods, and respondent populations) from 33 papers. We found a limited number of published papers on the evaluation of nonclinical fellowship programs, and most focused on outcomes associated with fellows or alumni. The most useful papers for our purposes clearly described the evaluation framework that guided the evaluation.

Keywords

Fellowship program; Program evaluation; Literature review; Service learning

1. Introduction

The public health workforce faces a number of challenges, including a small proportion of public health professionals with formal public health training (Lichtveld et al., 2001; Sellers et al., 2015), a changing public health landscape (Institute of Medicine Committee on Assuring the Health of the Public in the 21st Century, 2002) that necessitates

*Corresponding author. Present address: Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mailstop H16-3, Atlanta, GA 30333, USA. kvz3@cdc.gov (M. Paek).

Declaration of Competing Interest

None

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.evalprogplan.2020.101812.

resourcefulness and adaptation, and an aging workforce (Sellers et al., 2015). Comprehensive workforce development programs are needed in the field to address these challenges. Public health fellowships, which recruit qualified personnel, promote interdisciplinary work, and train the entering and current workforce for leadership and management (Council of State & Territorial Epidemiologists, 2004; Hayes, 2014), are an effective component of comprehensive workforce development (Koo & Miner, 2010). The Division of Scientific Education and Professional Development (DSEPD) in the Center for Surveillance, Epidemiology, and Laboratory Services (CSELS) at the U.S. Centers for Disease Control and Prevention (CDC) designs, implements, and evaluates fellowship programming to develop a competent, sustainable, and empowered public health workforce. Based on the concept of service learning (Furco, 1996; Wigington, Sobelson, Duncan, & Young, 2017), DSEPD fellowships engage public health professionals in experiential, on-the-job learning (Thacker, Dannenberg, & Hamilton, 2001). Each DSEPD fellowship is unique in its training discipline, duration, and targeting of educational level.

Government organizations in particular have increasingly called for data for decision-making and accountability (Centers for Disease Control & Prevention, 2012; United States Government Accountability Office, 2013). DSEPD ultimately should be able to show whether the investments in fellowship programs are meeting identified workforce development needs. DSEPD focused previous evaluation efforts on documenting program components (e.g., recruitment or curriculum) and implementation (e.g., participation or reach). Personnel used these data to make programmatic or operational decisions, but data on meaningful programmatic outcomes — especially for the portfolio of programs, rather than individual programs — remained limited. Outcomes data on the portfolio of programs would highlight areas of success and for improvement, informing future programmatic decisions and planning. Thus, DSEPD prepared to identify and then apply techniques to measure and share outcomes of its fellowship program portfolio. We started by looking to the literature for fellowship evaluation practices used. At minimum, we expected to find commonly used frameworks and methods to guide our evaluation of fellowships. At best, we hoped to find commonly used measures and indicators across fellowship evaluations that would help us to speak comprehensively about our own portfolio of fellowships.

This paper describes findings from our review of the literature on the evaluation of nonclinical fellowship programs. Our goal was to locate fellowship programs similar to our own and translate practices to our own context. However, what started as an exploratory, practical exercise in evaluation planning ended up revealing potential literature gaps that suggest opportunities for improvement in the fields of evaluation and workforce development. This paper does not describe a systematic review, but we believe our experience and observations could be valuable to other fellowship evaluators who might have a similar experience when searching for this type of literature. This paper describes the extent and nature of fellowship evaluations in the literature we found and considerations for workforce development professionals, particularly those tasked with the evaluation of fellowship programs.

2. Methods

Our review process involved 2 main steps as follows: first, search and screen; second, detailed review. The search and screen step took place over 2 phases. Phase I took place during 2015–2016 as a practical approach for the evaluation team to draw from for existing conceptual frameworks and practices. In February 2017, the evaluation team engaged a librarian for Phase II for a more thorough search to supplement our Phase I results. In the detailed review step, the analysts abstracted and coded information about program and evaluation characteristics described in each paper.

2.1. Search and screen, phase I

In Phase I of the search, 2 of the evaluators on the team, now referred to as analysts, independently searched both databases and search engines, including PubMed, ERIC, ProQuest, Google Scholar, Google, and the CDC Library's Primo Central Index. This process identified papers in both peer-reviewed journals and gray literature, published in English, and with no specific date ranges. Alternate locations of information, such as evaluations that could be accessed through an organization's website, were possible for identification during this phase if they matched our initial search terms. Search terms included combinations of the following: *fellowship, evaluation, assessment, workforce development, on the job training program, impact, success, outcomes, continuing education*, and *internship*. The search was inclusive of all sectors and industries. We added the following search terms, in conjunction with previously cited terms, to probe further into areas that we are aware have made concerted efforts to use workforce development training programs: *business, corporation, military*, and *Affordable Care Act*. Given the practical, iterative, and informal nature of Phase I, we did not document the total number of papers that resulted from the searches. Instead, we screened the titles and abstracts of our search results as we went along, by applying defined selection criteria (Appendix A in Supplementary material).

Our inclusion and exclusion criteria were largely based on definitions from our own fellowship programs, as the intention of our review was to identify practices that could translate to our own context. We specifically sought fellowships that, like ours, were experiential service programs that sought to develop the workforce in their field, rather than those that were part of a traditional educational program. We intentionally did not include evaluations of graduate medical training programs (eg, clinical residency) for such reasons. Graduate medical training programs are extensions of traditional medical education and required to practice within the profession. The need behind graduate medical training programs thus differs from that of DSEPD's fellowships. Consistent with our goals, we looked for evaluations that considered the whole program rather than a single component. Finally, we focused on selecting papers that described applied evaluations, not just conceptual frameworks for an evaluation, so we could identify common practices and lessons learned from a practical perspective.

We included papers that described evaluating programs meeting 2 core criteria, based on our definition of a fellowship: (1) primarily experiential (on-the-job) training, and (2) a

minimum duration of 1 month, with no work duties outside those associated with the fellowship experience.

We excluded papers according to criteria regarding the characteristics of the *program* and the *evaluation*. Associated with the program characteristics, we excluded papers about traditional educational programming, such as graduate medical education and other degree programs. We also excluded those focused on specific professional requirements or outcomes, such as those associated with clinical skills maintenance or continuing education credits. Associated with evaluation characteristics, we excluded papers focused on individual components of workforce development strategies (e.g., individual workshops or orientations). We also excluded those with a narrow scope, such as those focused on job satisfaction during training, or those that used a sole data collection method that we knew would be insufficient for our needs, such as pre- and posttests.

Using these selection criteria, the analysts separately screened their search results by assessing paper titles, abstracts, or executive summaries. The analysts also checked references of papers that met our selection criteria to identify additional fellowship program evaluations for inclusion in this review. Any references that appeared to meet selection criteria were put through the same screening. The 2 analysts then shared with each other their lists of papers for inclusion to verify the papers on each other's list met all criteria. For any papers on which there was disagreement, the analysts discussed and reached consensus on whether to include the paper.

2.2. Search and screen, phase II

Based on our results from Phase I, we consulted a CDC librarian to conduct a broad, independent search to ensure a thorough review of the peer-reviewed literature. Based on the experience of Phase I, we believed that a focus on peer-reviewed literature would offer more valuable findings and a greater breadth of comparable programs than a typical online search. The CDC librarian searched Medline (OVID; 1946–2017), PsychInfo (OVID; 1987–2017), and Scopus (without a specific date range), and adapted our search terms listed earlier. The librarian's search resulted in 479 unique papers (Appendix B in supplementary material). The analysts then used the same selection criteria and process described earlier to screen these 479 titles and abstracts. The goal remained to select papers that described programs similar to ours and evaluations that considered the program as a whole.

2.3. Detailed review

In the second step, each analyst independently conducted a more detailed review of half of the included papers and abstracted information about program characteristics (program description, sector, and length of the program) and the evaluation characteristics (primary evaluation type, framework for evaluation, data collection methods, and respondent populations) presented. All abstracted information was recorded as qualitative data in an Excel® (Microsoft, Inc., Seattle, WA) spreadsheet. During this detailed review, the analysts removed additional papers found to meet our exclusion criteria. In the end, Phase I produced 24 papers and Phase II produced 9 additional papers relevant to our review.

Using the abstracted data from the final set of included papers, the analysts grouped the qualitative data for each characteristic according to emergent themes and then developed codes to describe those themes. Codes were applied to papers for sector, length of the program, primary evaluation type, frameworks, data collection methods, and respondent populations. Because codes were not determined *a priori*, groupings were based on iterative discussion and analysis. The primary codes applied to the sector category were *education, health, library, public health, public service*, and *research*; secondary codes applied were *economics* and *leadership and management*. The length of the program was coded as < *one year, one year, two years, two and a half years*, and *five years*. Programs that did not have a set length of time for all fellows were coded as having *variable* duration and could exceed 5 years.

For the program's primary evaluation type, we used the following codes: *process*, *outcome*, and *impact*. We coded papers as *process* if the evaluation focused on assessing if and how well program activities and operations were implemented (Linnan & Steckler, 2002). Our *outcome* code was loosely defined and applied to papers that examined changes after the program (Frechtling, 2007). We defined *impact* evaluations to be those that sought to determine the extent of changes that could be attributed to the program (Frechtling, 2007). When an evaluation fit more than 1 code, a single code was applied based on the analysts' determination regarding the primary focus of the paper.

When coding for evaluation frameworks described in the papers, we used 3 primary codes: *Other*, *Program theory*, and *Not found*. We defined a framework as any structure explicitly articulated in the paper that described the concepts driving the evaluation inquiry. As we defined the term, frameworks could be based on external models or theories, or they could be based on the program's logic model or program theory. If the paper described external models or theories that guided their evaluation, we coded them as *Other*. If the paper described using their own program's logic model or program theory to guide their evaluation, we coded it as *Program theory* (Christie & Alkin, 2003). We coded any paper that presented no framework as *Not found*. For any papers that specified the model or theory they used, we noted that as open text.

The analysts coded data collection methods as survey, interviews, site visits, administrative data, observation, document review, focus groups, and other. Respondent populations were coded as alumni, fellows, supervisors (ie, of fellows), staff (ie, program staff who managed the fellowship program itself), host (ie, individuals who worked at the fellow's placement site, other than the supervisor), or other stakeholders.

3. Results

Thirty-three papers met our selection criteria for this review (Table 1). Of these, 27 were published in journals, including 23 in peer-reviewed journals, and 6 were from gray literature.

3.1. Program characteristics

Based on program descriptions included in the 33 papers, we found 28 unique programs or groups of programs across the 33 papers that met our criteria as fellowship programs, meaning that some papers described the same program or programs. Of these 28, 3 were about different groups of fellowship programs of a particular discipline. These disciplines were: postdoctoral programs (Davis, 2005), pharmacoeconomic or outcomes research fellowship programs (Maio & Lofland, 2004; Maio, Lofland, Doan et al., 2003; Maio, Lofland, & Nash, 2003), and residency programs for librarians (Brewer, 1998).

Nearly half (n = 15) of the 33 papers described programs in health-related sectors (not focused exclusively on training in research), with foci including public health, health economics, and leadership and management in public health and healthcare settings. Fourteen of the 33 papers described research training programs, including health research (eg, biomedical research) programs and programs associated with scientific research. The 4 remaining papers described fellowships for teaching, library science, and public policy and service.

Program length varied widely, from 6 weeks to 5 years or more, but most papers (n = 19) described programs that lasted from 1 to 2 years.

3.2. Evaluation characteristics

Looking at characteristics of the evaluations, we found most of the papers to describe outcome evaluations (n = 25). Four papers were coded as impact evaluations, and an additional 4 as process evaluations. Twenty-five papers did not indicate use of a framework to guide the evaluation. Eight papers did indicate clear use of a framework; 6 of the these used program theory and 2 used other frameworks, specifically the Donebedian model (Maio, Lofland, Nash, 2003) and Bloom’s Taxonomy (Maio & Lofland, 2004).

Overall, surveys were the most common data collection method described (n = 24), with administrative data (n = 14; e.g., applications, progress reports, and curricula vitarum) and interviews (n = 12) the next most common. Other data collection methods described were observation (n = 4), document review (n = 3), site visits (n = 2), and focus groups (n = 1). Papers were approximately evenly split between those using some combination of data collection methods (n = 18) and those using a single data collection method (n = 15). Among evaluations with only 1 data collection method, 9 used surveys alone, 5 relied on administrative data, and 1 relied on interviews.

Alumni and current fellows were the most common respondent populations (n = 23 and n = 18, respectively). Other common respondent populations included supervisors and mentors (n = 9), fellowship program staff (n = 5), and other individuals who worked with the fellow at the host site (n = 5).

4. Discussion

Our review led us to 3 main observations. Our first observation is that most papers did not describe frameworks that guided their evaluation, but of those that did, those that described

using program theory were the most useful. This was because these evaluations explained *why* they were looking at certain outcomes or variables and their significance to the program. The explanation of how the results connected to the intended goals of the program added value to the interpretation and understanding of the evaluation. For those without a clear program description or evaluation framework to ground the purpose and direction of the evaluation, we found the data presented within these papers to be largely descriptive and not clearly indicative of program success or achievement. This challenged our ability to translate concrete practices for our own work.

As a result, we found program theory to be beneficial in structuring an evaluation (Christie & Alkin, 2003), and want to emphasize the value of using and sharing program theory. A clear program theory contributes to program planning by clarifying program activities and related goals, and also provides essential meaning to its evaluation. Especially given the complexity of fellowship programs and the absence of a well-established fellowship evaluation framework found in the literature, we consider this a promising approach for evaluating fellowship programs. For evaluations that already use program theory, we recommend that authors explain this use in their reports or publications; this could help add context and meaning to the presentation of methods and data for external audiences. For those seeking to ground their upcoming evaluation in a framework, we find the program theory approach to be practical and easily adaptable to a variety of setting and programs.

Our next observation is that across the papers, the respondent groups were primarily alumni or current fellows. From our own experience, we understand that these groups are likely to be most accessible as respondents, and that these groups might be viewed as the direct beneficiaries of the fellowship. Our fellowships, however, are designed around the concept of service-learning (Furco, 1996), which intends to benefit both the entity receiving services *and* the fellow. We would be interested in seeing how future evaluators and authors incorporate an approach that collects data from a variety of different perspectives, to understand better the different roles and mechanisms that are critical to a successful fellowship program.

Our final observation is that our results included fewer papers than anticipated. Having conducted a literature search twice, first by the analysts, and second by a librarian, the final 33 papers did not quite meet our expectations in terms of volume. Although we recognize our selection criteria were limiting, we expected a larger number of final papers for review. Further, after completing this analysis, we became aware of a high-profile fellowship evaluation paper (Jones, MacDonald, Volkov, & Herrera-Guibert, 2014) that met our criteria yet did not appear in our search results. The fact that our searches found few relevant papers *and* did not find some relevant papers suggests (1) a potential gap in the literature, and (2) an issue with effective accessibility of relevant literature. Should a potential gap in the literature exist, we encourage our peers to share their work to guide best practices and lessons learned. Dissemination is key to allow others to learn from previously conducted evaluations. However, dissemination alone is not enough. The field can only expand and improve with papers that are retrieved when needed. Making internal reports publicly available on organization's websites would increase the availability of fellowship evaluations in the gray literature. We recognize that this is likely already occurring to some extent; however, that the

gray literature is be published in such disparate ways can make retrieval challenging for most practitioners. We recommend publication in journals associated with effective indexing to increase the likelihood that the publication comes up in database search results.

5. Limitations

This effort evolved out of a practical need for useful guidance for evaluating fellowships. As a result, our methods were informal prior to consulting a CDC librarian and are likely not replicable. Specific searches for existing literature reviews on this topic could have increased the number of papers that met our needs. One such literature review, for example, found a lack of outcomes data in the published literature on postgraduate biomedical and public health training programs (Faupel-Badger, Nelson, Marcus, Kudura, & Ngheim, 2013). Additionally, further information could have been found through organization websites and other alternate sources had we chosen to pursue a review of specific workforce development programs. We continued to adapt the process before engaging a librarian to investigate the literature base more thoroughly. However, because of our findings we felt compelled to share our experience and our recommendations, as we believe they are of interest to other fellowship evaluators and potentially applicable to the broader field of public health workforce development.

We recognize that our set of criteria for including papers in our review was narrow, which limited our scope. For example, by excluding papers describing evaluations that used a single method, such as relying solely on pre- and posttests, we limited our set of results. However, the intent of our search was to find frameworks and methods we could apply to our service-learning fellowships. From the beginning, we knew that a pre- and posttest would not be sufficient to meet our needs and were invested in finding practices that evaluated the fellowship more comprehensively. We also recognize that our set of criteria was dependent upon a paper's voluntary inclusion of program description. Under these guidelines, the analysts eliminated at least 30 promising papers during the screening step, when we could not discern whether the program matched our criteria. Although these criteria resulted in narrowing the scope of our search, using these criteria revealed that program descriptions are frequently underrepresented, and brought us to one of our recommendations.

Finally, conducting searches solely through online venues and from CDC computers might have biased our search results toward health-related topics. However, it is possible that fellowships are more common in these sectors, or that these sectors are more likely to evaluate and disseminate their findings. Other sectors might also use terminology other than *fellowship*; perhaps a term like *apprenticeship* might be useful in future literature reviews. The low volume of papers identified from outside the health and science sectors only emphasizes the need for literature that is more available and accessible for the common evaluation practitioner.

6. Lessons learned

Our review highlights some areas that could be useful for fellowship evaluators' and program planners' future work. The intention behind the associated recommendations is to increase the rigor in the field of fellowship evaluation and support fellowship planning. We also had several lessons learned that will inform our future practice and may also be useful to readers.

First, our experience with this review leads us to encourage fellowship evaluators to share their work and build the body of literature on fellowship evaluations. Sharing approaches and knowledge will only help to advance our individual work and the field as a whole. Although evaluations of CDC fellowship programs were included in the final list of included papers, their relatively low number reinforced our original program goals to increase evaluation among our own fellowship programs. We will learn from the experiences presented in the literature, and contribute our findings to continue to expand it.

Second, as stated earlier, we found our selection criteria to be limiting. Our expectation was that we would find ample content in the literature to assist in the design of our fellowship evaluation projects. While our search terms indeed produced a large number of results, many were so narrow in focus, we found little value in them; this was the reason behind using the intentionally restrictive selection criteria. We elected not to expand the selection criteria or search terms, under the assumption that additional potentially useful articles would not yield adequate findings to warrant the additional time in reviewing. However, it is certainly possible that relevant and useful information, albeit more resource intensive to glean, could have been found had we expanded our scope.

In a final and related point, in terms of our methodology and experience conducting this review, a lesson learned for us is the potential value of pivoting to a more systematic approach earlier in our process, especially once we started to realize that the literature was not as robust as we had hoped. Following a systematic review process would not necessarily have contributed additional evidence in support of our initial goal of identifying evidence-based fellowship evaluation methods. It could have, however, provided a more comprehensive review and synthesis of the literature, and as a result more specific and concrete recommendations for the field of fellowship evaluation.

7. Conclusion

The field of public health is at a turning point that requires a robust workforce prepared to meet emerging needs, and fellowship programs can play an important role in developing this workforce (Council of State & Territorial Epidemiologists, 2004; Hayes, 2014). Fellowship staff should rely on evidence-based methods to produce data for action, to ensure effective fellowship programs that contribute to developing a robust public health workforce. We began this exercise as evaluation practitioners, exploring the literature for guidance to apply to our own work. We found a limited number of published papers on fellowship evaluation, that focused on a subset of the potential beneficiaries, and which largely do not sufficiently describe their programs or the evaluation frameworks guiding the evaluation. However, the

papers that clearly delineated how their evaluation tied directly to the program were most useful, and resulted in us turning to program theory to guide our work. Our call is for increased use of program theory and evaluation frameworks within a larger volume of fellowship evaluation literature to improve the literature base on which fellowship staff and evaluators can rely.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by an appointment to the Research Participation Program at the Centers for Disease Control and Prevention administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and CDC. We thank Joanna Taliano of the Centers for Disease Control and Prevention, Library Science Branch, for assistance with identifying relevant literature.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Biography

Margaret Paek, MPH, is a program evaluator with Northrop Grumman, currently working with CDC's Division of Healthcare Quality Promotion. Previously, she completed a two-year appointment to the Oak Ridge Institute for Science and Education's Research Participation Program as an evaluation fellow in CDC's Division of Scientific Education and Professional Development. Ms. Paek has worked in program evaluation since 2013, and has provided evaluation support to nonprofit organizations, foundations, academic institutions, and CDC.

Connor Radkey, MDP, is a monitoring and evaluation manager with United Methodist Committee on Relief. He previously completed a three-year appointment with the Oak Ridge Institute for Science and Education's Research Participation Program as an evaluation fellow in CDC's Division of Scientific Education and Professional Development. Mr. Radkey has been an evaluator for the past six years, having worked with international non-governmental organizations, nonprofits, and CDC to design and implement monitoring and evaluation activities.

Sally Honeycutt, MPH, is the Evaluation Lead for CDC's Division of Scientific Education and Professional Development. As a professional with over two decades' experience in public health, Ms. Honeycutt has worked as a program evaluator for the past 13 years. In collaboration with program stakeholders, she has developed and implemented evaluation plans for initiatives and programs conducted by state and local health departments, community-based organizations, academic institutions, and CDC.

M. Kathleen Glynn, DVM, MPVM, is the Deputy Director for Science of CDC's Division of Scientific Education and Professional Development. She has been at the CDC for over 20 years, contributing to the agency's goals in both disease-specific areas such as HIV/AIDS

and zoonotic diseases, and cross-cutting programs such as surveillance. For the past five years she has focused on advancing the science of public health workforce development.

References

Acker AL, Freeman JD, & Williams DM (1988). A medical school fellowship program for minority high school students. Journal of Medical Education, 63(3), 171–175. [PubMed: 3346892]

Beecroft PC, Kunzman L, & Krozek C (2001). RN internship: outcomes of a one-year pilot program. The Journal of Nursing Administration, 31(12), 575–582. [PubMed: 11771462]

Brewer J (1998). Post-master's residency programs: Enhancing the development of new professionals and minority recruitment in academic and research libraries. College and Research Libraries, 59(6), 528–537. 10.5860/crl.59.6.528.

Bruheim M, Woods KV, Smeland S, & Nortvedt MW (2014). An educational program to transition oncology nurses at the Norwegian Radium Hospital to an evidence-based practice model: Development, implementation, and preliminary outcomes. Journal of Cancer Education, 29(2), 224–232. 10.1007/s13187-013-0575-9. [PubMed: 24197689]

Buffington J, Bellamy PR, & Dannenberg AL (1999). An elective rotation in applied epidemiology with the Centers for Disease Control and Prevention (CDC), 1975–1997. American Journal of Preventive Medicine, 16(4), 335–340. [PubMed: 10493292]

Centers for Disease Control and Prevention (2012). Improving the use of program evaluation for maximum health impact: Guidelines and recommendations. Retrieved from https://www.cdc.gov/eval/materials/finalcdcevaluationrecommendations_formatted_120412.pdf.

Christie CA, & Alkin MC (2003). The user-oriented evaluator's role in formulating a program theory: Using a theory-driven approach. The American Journal of Evaluation, 24(3), 373–385. 10.1016/S1098-2140(03)00054-7.

Council of State and Territorial Epidemiologists (2004). CSTE special report: Workforce development initiativeRetrieved fromhttp://www.cste2.org/webpdfs/Workforcesummit.pdf.

Creasy RK (1996). Evaluation of postdoctoral physician-scientist training programs in obstetrics and gynecology. American Journal of Obstetrics and Gynecology, 175(2), 494–495. [PubMed: 8765275]

Davis G (2005). Doctors without orders. American Scientist, 93(Suppl. 3).

Dick VR, Masters AE, McConnon PJ, Engel JP, Underwood VN, & Harrison RJ (2014). The CDC/Council of State and Territorial Epidemiologists Applied Epidemiology Fellowship Program: Evaluation of the first 9 years. American Journal of Preventive Medicine, 47, S376–S382. 10.1016/j.amepre.2014.07.022 (5, Supplement 3). [PubMed: 25439260]

Dores GM, Chang S, Berger VW, Perkins SN, Hursting SD, & Weed DL (2006). Evaluating research training outcomes: Experience from the cancer prevention fellowship program at the National Cancer institute. Academic Medicine, 81(6), 535–541. [PubMed: 16728801]

Dudley DJ (1997). Optimism for perilous times: A survey of American Association of Obstetricians and Gynecologists Foundation and Reproductive Scientist Development Program postdoctoral research fellows. American Journal of Obstetrics and Gynecology, 176(4), 814–818. [PubMed: 9125605]

Faupel-Badger JM, Nelson DE, Marcus S, Kudura A, & Ngheim E (2013). Evaluating postgraduate public health and biomedical training program outcomes: Lost opportunities and renewed interest. Journal of Cancer Education, 28(1), 18–26. [PubMed: 23225110]

Faupel-Badger JM, Raue K, Nelson DE, & Tsakraklides S (2015). Alumni perspectives on career preparation during a postdoctoral training program: A qualitative study. CBE-Life Sciences Education, 14(1), 10.1187/cbe.14-06-0102.

Feeley FR, Connelly P, McCoy K, Richards SC, & Vian T (2006). Evaluation of the Pfizer Global Health Fellows ProgramRetrieved fromhttp://www.bu.edu/cghd/files/2010/10/FINAL-GHF-evaluation-report-short.pdf.

Frechtling JA (2007). Logic modeling methods in program evaluation (1st ed.). San Francisco, CA: Jossey-Bass.

Furco A (1996). Service-learning: A balanced approach to experiential education Expanding boundaries: Serving and learning. Washington, D.C: The Corporation for National Service.

Hall DE, & Marshall RL (2006). Evaluation of a 16-week critical care internship program using a staff development program effectiveness evaluation tool. Journal for Nurses in Staff Development, 22(3), 134–143. [PubMed: 16760774]

Hayes D (2014). Insights in public health: Strengthening the epidemiology workforce through mentorship: Practicum and fellowship experiences in the family health services division at the Hawaiì department of health. Hawai'i Journal of Medicine & Public Health: A Journal of Asia Pacific Medicine & Public Health, 73(3), 94–97.

Institute of Medicine Committee on Assuring the Health of the Public in the 21st Century (2002). In the future of the public's health in the 21st century. Washington (DC): National Academies Press (US).

Jones SR, & Filerman GL (2007). An analysis of the impact of the David A. Winston health policy fellowship. The Journal of Health Administration Education, 24(1), 33–41. [PubMed: 17847865]

Jones D, MacDonald G, Volkov B, & Herrera-Guibert D (2014). Multisite evaluation of field epidemiology training programs: Findings and recommendations. Atlanta, Georgia: Centers for Disease Control and Prevention 2014.

Käser M, Maure C, Halpaap BMM, Vahedi M, Yamaka S, Launois P, & Casamitjana N (2016). Research capacity strengthening in low and middle income countries – An evaluation of the WHO/TDR career development fellowship programme. PLoS Neglected Tropical Diseases, 10(5), e0004631 10.1371/journal.pntd.0004631. [PubMed: 27223888]

Koo D, & Miner K (2010). Outcome-based workforce development and education in public health. Annual Review of Public Health, 31(1), 253–269. 10.1146/annurev.publhealth.012809.103705.

Lichtveld MY, Cioffi JP, Baker EL Jr., Bailey SB, Gebbie K, Henderson JV, … Miner K (2001). Partnership for front-line success: A call for a national action agenda on workforce development. Journal of Public Health Management and Practice, 7(4).

Linnan L, & Steckler A (2002). Process evaluation for public health interventions and research. San Francisco, CA: Jossey-Bass.

MacAllum K, & Gallup-Black A (2003). Toward diversity in public service: A report to the Ford Foundation on the public policy and international affairs (PPIA) fellowship program 1980–2000Retrieved from.

Macro International Inc (2000). Evaluation of the Hubert H. Humphrey fellowship program. Retrieved from.

Maio V, & Lofland JH (2004). Effectiveness of pharmacoeconomic and outcomes research fellowship programs within the United States. American Journal of Pharmaceutical Education, 68(1–5), YYY1–YYY8.

Maio V, Lofland JH, Doan Q, Duttagupta S, Khan ZM, Phillips AL, … Yokoyama K (2003). Organizational structure and educational components of pharmacoeconomic fellowship programs: A fellows' perspective. American Journal of Pharmaceutical Education, 67(4), 105 10.5688/aj6704105.

Maio V, Lofland JH, & Nash DB (2003). Pharmacoeconomic fellowships: Preceptors' views regarding adherence to the American College of Clinical Pharmacy guidelines. American Journal of Pharmaceutical Education, 67(3), 75 10.5688/aj670375.

Martinez A, Epstein C, Parsad A, & Whittaker K (2012). Evaluation of NSF's international research fellowship program: Final reportRetrieved from Bethesda, MD.

Meyer GS, Edwards JN, & Blumenthal D (1994). Experience of the Robert Wood Johnson health policy fellowship. Health Affairs, 13(2), 264.

Mihaly K, Master B, & Yoon C (2015). Examining the early impacts of the leading educators fellowship on student achievement and teacher retention. Retrieved from.

Pincus SH, Wolff EM, & Melander EH (2002). The VA National quality scholars fellowship program: Current status, future directions. Quality Management in Health Care, 10(3), 71–76. [PubMed: 12512468]

Pion GM, & Hammond CB (2005). The American Association of Obstetricians and Gynecologists Foundation Scholars Program: Additional data on research-related outcomes. American Journal of Obstetrics and Gynecology, 193(5), 1733–1739. 10.1016/j.ajog.2005.07.052. [PubMed: 16260218]

Pion G, & Ionescu-Pioggia M (2003). Bridging postdoctoral training and a faculty position: Initial outcomes of the Burroughs Wellcome Fund Career Awards in the Biomedical Sciences. Academic Medicine, 78(2), 177–186. [PubMed: 12584096]

Sellers K, Leider JP, Harper E, Castrucci BC, Bharthapudi K, Liss-Levinson R, … Hunter EL (2015). The public health workforce interests and needs survey: The first national survey of state health agency employees. Journal of Public Health Management and Practice, 21, S13–S27. 10.1097/phh.0000000000000331. [PubMed: 26422482]

Sopher CJ, Adamson BJS, Andrasik MP, Flood DM, Wakefield SF, Stoff DM, … Fuchs JD (2015). Enhancing diversity in the public health research workforce: The research and mentorship program for future HIV vaccine scientists. American Journal of Public Health, 105(4), 823 10.2105/AJPH.2014.302076. [PubMed: 25122028]

Steiner JF, Lanphear BP, Curtis P, & Vu KO (2002). The training and career paths of fellows in the National Research Service Award (NRSA) Program for Research in Primary Medical Care. Academic Medicine, 77(7), 712–718. [PubMed: 12114148]

Stoll L, Swanwick T, Foster-Turner J, & Moss F (2011). Leadership development for junior doctors: What can we learn from “Darzi” Fellowships in clinical leadership? The International Journal of Leadership in Public Services, 7(4), 273–286. 10.1108/17479881111194161.

Thacker SB, Dannenberg AL, & Hamilton DH (2001). Epidemic Intelligence Service of the Centers for Disease Control and Prevention: 50 years of training and service in applied epidemiology. American Journal of Epidemiology, 154(11), 985–992. 10.1093/aje/154.11.985. [PubMed: 11724713]

United States Government Accountability Office (2013). Program evaluation: Strategies to facilitate agencies’ use of evaluation in program management and policy making. Retrieved from Washington, D.C.:http://www.gao.gov/products/GAO-13-570.

Vian T, Koseki S, Feeley FG, & Beard J (2013). Strengthening capacity for AIDS vaccine research: Analysis of the Pfizer global health fellows program and the international AIDS vaccine initiative. BMC Health Services Research, 13, 378 https://doi.org/10.1007/s11266-008-9072-xhttps://doi.org/10.1186/1472-6963-13-378https://doi.org/10.1186/1472-6963-13-378 . [PubMed: 24088300]

Walmsley J, Hockey P, Kitsell F, & Sewell A (2012). Improving Global Health: a win-win for leadership development and a fairer world? The International Journal of Leadership in Public Services, 8(4), 216–231. 10.1108/17479881211323607.

Wigington CJ, Sobelson RK, Duncan HL, & Young AC (2017). Service learning in public health: Exploring the benefit to host agencies in CDC’s public health associate program. Journal of Public Health Management and Practice : JPHMP. 10.1097/phh.0000000000000523.

Woltring C, Constantine W, & Schwarte L (2003). Does leadership training make a difference? The CDC/UC public health leadership institute: 1991–1999. Journal of Public Health Management and Practice, 9(2), 103–122. [PubMed: 12629912]

Table 1

Inventory of included papers and coding results.

Paper	Sector	Length of program	Primary evaluation type	Framework for evaluation	Data collection methods	Respondent population(s)
Acker, Freeman, & Williams, 1988	Research	< 1 year	Outcome	Not found	Survey	Alumni
Beecroft, Kunzman, & Krozek, 2001	Health	1 year	Outcome	Program theory	Survey, administrative data	Fellows, supervisors
Brewer, 1998	Library	1 year	Outcome	Not found	Survey	Alumni
Bruheim, Woods, Smeland, & Nortvedt, 2014	Health	Variable	Outcome	Not found	Survey, observations	Fellows, staff
Buffington, Bellamy, & Dannenberg, 1999	Public health	< 1 year	Outcome	Not found	Administrative data	Alumni
Creasy, 1996	Research	2 years	Outcome	Not found	Administrative data	Fellows, alumni
Davis, 2005	Research	Variable	Process	Not found	Survey	Fellows
Dick et al., 2014	Public health	2 years	Outcome	Not found	Survey, administrative data	Alumni, supervisors
Dores et al., 2006	Research	2.5 years	Outcome	Not found	Administrative data	Alumni
Dudley, 1997	Research	2 years	Outcome	Not found	Survey	Fellows, alumni
Faupel-Badger, Raue, Nelson, & Tsakraklides, 2015	Research	2.5 years	Process	Not found	Interviews	Alumni
Feeley, Connelly, McCoy, Richards, & Vian, 2006	Research	< one year	Outcome	Not found	Survey, interviews	Fellows, alumni, supervisors, staff
Hall & Marshall, 2006	Health	< 1 year	Outcome	Program theory	Observation, administrative data, survey	Fellows, supervisors
Jones & Filerman, 2007	Health	1 year	Outcome	Not found	Survey, interviews	Alumni
Käser et al., 2016	Research	1 year	Outcome	Program theory	Document review, survey, interviews, focus groups	Fellows, host, staff, alumni, other stakeholders
MacAllum & Gallup-Black, 2003	Public service	< 1 year	Outcome	Not found	Survey, interviews, administrative data	Fellows, alumni, host
Macro International Inc., 2000	Public service - leadership & management	1 year	Outcome	Not found	Survey, interviews, site visits, administrative data	Alumni
Maio & Lofland, 2004	Health - economics	2 years	Outcome	Other: Bloom's Taxonomy of Educational Objectives	Survey	Alumni
Maio, Lofland, Doan et al., 2003	Health - economics	2 years	Process	Other: Donabedian's framework of structure, process, and outcomes	Survey	Supervisors
Maio, Lofland, Doan, et al., 2003	Health - economics	2 years	Process	Not found	Survey	Fellows, alumni

Paper	Sector	Length of program	Primary evaluation type	Framework for evaluation	Data collection methods	Respondent population(s)
Martinez, Epstein, Parsad, & Whittaker, 2012	Research	Variable	Impact	Not found	Survey, administrative data	Fellows, supervisors
Meyer, Edwards, & Blumenthal, 1994	Health	1 year	Impact	Not found	Survey	Fellows, alumni
Mihaly, Master, & Yoon, 2015	Education — leadership & management	2 years	Impact	Not found	Other, administrative data	Fellows, alumni
Pincus, Wolff, & Melander, 2002	Health - leadership & management	2 years	Outcome	Not found	Site visits, interviews	Fellows, host
Pion & Ionescu-Pioggia, 2003	Research	5 years	Outcome	Program theory	Survey, administrative data	Fellows, alumni
Pion & Hammond, 2005	Research	Variable	Outcome	Not found	Administrative data, other (supplemental web search)	Alumni
Sopher et al., 2015	Research	Variable	Outcome	Not found	Survey, interviews	Fellows, supervisors
Steiner, Lanphear, Curtis, & Vu, 2002	Research	Variable	Outcome	Not found	Survey	Alumni
Stoll, Swanwick, Foster-Turner, & Moss, 2011	Health - leadership & management	1 year	Outcome	Program theory; drawn from Kirkpatrick, Guskey, Stufflebeam	Interviews, survey, document review, observation	Fellows, host, supervisors
Thacker et al., 2001	Public health	2 years	Outcome	Not found	Not specified	Alumni
Vian, Koseki, Feeley, & Beard, 2013	Research	< 1 year	Outcome	Not found	Administrative data, interviews	Fellows, host
Walmsley, Hockey, Kitsell, & Sewell, 2012	Health - Leadership & management	1 year	Outcome	Not found	Document review, survey, interviews, observation	Alumni, supervisors, staff
Woltring, Constantine, & Schwarte, 2003	Public health - Leadership & management	1 year	Impact	Program theory	Survey, interviews	Alumni, staff, other stakeholders