

Comparing and Contrasting a Program versus System Approach to Evaluation: The Example of a Cardiac Care System

Ralph Renger

Just Evaluation Services, LLC

Jessica Renger, Stewart I. Donaldson

Claremont Graduate University

Jirina Renger

Walden University

Gary Hart

University of North Dakota

Andrew Hawkins

ARTD Consultants

Abstract: *This article examines the differences between approaching an evaluation problem from a program perspective and doing so from a systems perspective. The terms program, systems, systems thinking, and systems concepts are first defined. Then, using an actual evaluation of a cardiac care system, it is shown how initial investments in a program theory approach were deemed inadequate to account for the influence of external factors on patient outcomes. It was decided that a systems thinking approach was more appropriate for evaluating the interactions between several agencies comprising the cardiac care system. It is then shown how System Evaluation Theory (SET) was used to systematically apply different systems concepts to define and evaluate the cardiac care system. The discussion compares and contrasts the program and system evaluation approaches, noting the conditions under which each is more appropriate. It concludes by noting scope and cost differences between the two approaches.*

Keywords: *evaluation, systems, systems thinking*

Résumé : *Le présent article examine les différences entre l'utilisation d'une perspective de programme ou d'une perspective systémique lors d'une d'évaluation. Tout d'abord, on définit les termes programme, systèmes, théorie des systèmes et concept des systèmes. Puis, en utilisant l'évaluation réelle d'un système de soins cardiaques, on montre comment les investissements initiaux, dans une approche de*

Corresponding author: Ralph Renger, Just Evaluation Services, LLC, 14777 E. Circle M Ranch Place, Vail, AZ 85641; ralph@justevaluation.com

théorie des programmes, ont été jugés inadéquats pour tenir compte de l'influence des facteurs externes sur les résultats cliniques des patients. Il a été décidé que l'approche de théorie des systèmes était plus appropriée pour l'évaluation des interactions entre plusieurs agences formant le système de soins cardiaques. On montre ensuite que la théorie de l'évaluation des systèmes a été utilisée pour appliquer systématiquement différents concepts de systèmes en vue de définir et d'évaluer le système de soins cardiaques. La discussion compare et oppose les approches d'évaluation de systèmes et de programme, notamment les conditions les plus appropriées à chacune des approches. L'article se conclut en notant les différences en matière de portée et de coût des deux approches.

Mots clés : évaluation, systèmes, théorie des systèmes

Systems thinking as a way to improve evaluations continues to garner much attention in the evaluation literature (AEA, 2018; Hummelbrunner, 2011; Renger, 2015; Renger, Foltysova, Ienuso, Renger, & Booze, 2017; Williams & Britt, 2014; Williams & Hummelbrunner, 2010). The AEA Systems in Evaluation Topical Interest Group defines systems thinking as “a way of thinking based on core systems concepts” (AEA, 2018, p. 6). Systems concepts, in turn, “are those that have come to define the systems field” (AEA, 2018, p. 6). Examples of systems concepts used in evaluations include, but are not limited to, boundaries, elements, perspectives, feedback loops, cascading failures, surges, reworks (waste), and reflex arcs (Renger, 2016; Renger, Foltysova, Ienuso, et al., 2017; Renger, Keogh, Hawkins, Foltysova, & Souvannasacd, 2018; Ulrich, 2002).

Our review of the evaluation literature suggests that systems concepts are being applied to evaluate two related, but distinct, evaluands: programs and systems. The failure to delineate between these two evaluands has created significant confusion in the evaluation community (Renger, Wood, Williamson, & Krapp, 2011). Therefore, for our discussion purposes we define a program as a “single, specific purpose/activity/intervention and is ancillary to the main function of the organization. Typically, the longevity and funding of a program are subject to internal and external factors” (Pima County Health Department, 2018). Further, we adopt the AEA (2018, p. 6) definition of a system as “a set of interrelated elements that interact to achieve an inherent or ascribed purpose” (Ackoff, 1971; Meadows & Wright, 2008).

Using these definitions, it is reasonable to extrapolate that a program has narrower boundaries than a system, and multiple programs operating together can constitute elements of a system (Renger, Atkinson, Renger, Renger, & Hart, 2019). Further, a program is assumed to have a finite duration, whereas a system is enduring. While the concepts are distinct, they are intertwined: Programs represent deliberate interventions into existing systems. Any systems evaluation that leads to change will involve one or more interventions. The difference for evaluation is whether the orientation of the evaluator is first and foremost to understanding the value of an intervention or the value (or efficiency and effectiveness) of a given system. A systems evaluation would often provide the rationale by which

to select an intervention that may be appropriate to correct any current deficiencies in a current system. In this way a systems evaluation can operate like a needs analysis. Its orientation is to identify what is needed in this particular system at this particular point in time. This is different from mainstream program evaluation, which will often seek to warrant the enduring value of an intervention. As we discuss below, the treatment of “context” is key. The relative emphasis on the context for an intervention as something to be controlled or something to be embraced often serves to separate approaches to evaluation, including program and systems evaluation.

It is the authors’ collective experience, based on workshop participant feedback, student feedback, and participation in numerous international evaluation panels, that many evaluators struggle to concretely understand the difference between applying systems thinking and systems concepts to evaluating programs versus systems (GIZ, 2011). The authors receive numerous requests for case examples to illustrate how an evaluation would differ if approached from a program versus a systems perspective. The purpose of this article is to meet this need by using a cardiac arrest intervention evaluation case study. We begin with an overview of the case study, move on to describe how the intervention was evaluated using a program evaluation approach, and conclude by demonstrating how the scope and questions shifted when evaluating the intervention using a systems approach.

BACKGROUND TO THE CASE STUDY

As defined by the National Institutes of Health (NIH, 2019), “Sudden cardiac arrest [SCA] is a condition where the heart suddenly and unexpectedly stops beating.” When the heart stops beating, blood stops flowing. Cardiopulmonary resuscitation (CPR) is critical to keep the brain oxygenated when the heart stops beating (Bohm, Rosenqvist, Herlitz, Hollenberg, & Svensson, 2007). Bystanders and/or emergency medical professionals who are able to deliver high-quality CPR are a key factor in determining whether someone survives an SCA (Guzy, Pearce, & Greenfield, 1983). However, delivering high-quality CPR is exhausting and difficult to sustain (Meaney et al., 2013). Therefore, to ensure that high-quality compressions are sustained, first responders will often take turns providing CPR. However, the research shows that even small pauses that occur when alternating to provide CPR can decrease the likelihood of good patient outcomes (Sutton et al., 2009).

The difficulty associated with providing high-quality CPR is exacerbated in a rural emergency response. In rural areas, transport times from the scene to the nearest hospital often exceed 30 minutes, making it almost impossible to sustain high-quality CPR. Further complicating the problem is that rural ambulance services are often understaffed. As a result, often there are not enough EMTs available on a response call to alternate CPR.

In response to this problem, the Leona M. & Harry B. Charitable Trust invested several million dollars in providing rural EMS agencies, critical access

hospitals (CAHs) (small rural hospitals with fewer than 25 beds), and tertiary-care facilities across seven Midwest and Mountain-west states with the LUCAS™2 ([Helmsley, 2019](#)). The LUCAS™2 is a battery-operated mechanical CPR device that provides high-quality chest compressions for several hours. The funder approached the lead author asking for assistance in evaluating the extent to which the LUCAS™2 device led to better patient outcomes, defined as improved survival rates and neurological outcomes ([Sandroni & Geocadin, 2016](#)).

PROGRAM EVALUATION APPROACH

We (our evaluation team) initially reasoned that the evaluation of the LUCAS™2 intervention required both a program improvement (i.e., process) and merit and worth (i.e., outcome) evaluation focus ([Mark, Henry, & Julnes, 2000](#)). The process evaluation focused, first, on whether the LUCAS™2 training was delivered with fidelity, and, second, on whether the LUCAS™2 was being properly deployed in the field.

With respect to the first process objective, an evaluation team member attended all LUCAS™2 training sessions and monitored the extent to which the training was standardized across different trainers. One recommendation arising from the process evaluation was to revisit the train-the-trainer model, since training quality deteriorated as it was passed on from trainer to trainer. A second recommendation was to create a detailed LUCAS™2 training manual to improve training standardization (see [Physio-control, 2016](#)).

With respect to the second process objective, we collected data on LUCAS™2 deployments using a call center. EMS services were required to contact our call center after each LUCAS™2 deployment. An important finding arising from this process evaluation objective was that the stabilization strap, essential to prevent device migration producing unintended organ damage, was not consistently being secured. The process evaluation revealed that under periods of high stress, as is the case in a time-urgent SCA event, some EMTs were forgetting to use the stabilization strap. The training was then modified to better simulate stress by adding a timed component.

To evaluate impact, we opted to use a theory-driven evaluation (TDE) approach ([Chen, 1990](#); [Donaldson, 2003, 2007](#)). The use of program theory allows evaluators to identify the relevant immediate, intermediate, and long-term outcomes needing to be evaluated ([Renger & Titcomb, 2002](#)). It also allows verification as to whether proposed activities are appropriately targeted to produce the intended change. Finally, program theory allows evaluators to examine whether the measurement tools to capture that change are appropriate ([Renger & Titcomb, 2002](#)).

The theory underpinning the LUCAS™2 intervention was relatively straightforward and grounded in replicable and valid research ([Eisenberg, 2013](#)). The logic model in [Table 1](#) summarizes the key intervention elements. The program theory underlying the LUCAS™2 states that if high-quality CPR is provided until

Table 1. LUCAS™2 logic model

| Intervention assumptions | Intervention | Inputs | Outputs | Outcomes |
|-----------------------------------------------------------------|--------------|------------------|----------------------|---------------------------------------------------------------------------------------|
| Delivering high-quality CPR will improve patient survival rates | LUCAS™2 | Training manuals | # trained # deployed | Increased blood flow Improved neurological functioning Increased survival rates |

the patient reaches definitive care, then there is an increased likelihood of survival and better patient outcomes.

The impact evaluation was focused on answering whether deploying the LUCAS™2 resulted in better patient outcomes. Two primary outcomes were evaluated: patient survival rates and patient neurological outcomes. We employed a mixed methods approach for the impact evaluation by cross-referencing the EMT's self-report with the hospital database.

The impact evaluation results revealed that there were many SCAs where the LUCAS™2 was deployed but the patient did not survive. We learned through our call-center and focus-group interviews that there were many factors, not included in our program theory, that influenced the final patient outcomes. For example, we learned that in some SCA cases the LUCAS™2 was deployed on a patient who was dead on arrival. The EMTs deployed the LUCAS™2 to comfort family members so that they would know that something was done to help their loved ones. We also learned that in some cases the emergency (911) dispatcher did not indicate that the emergency was an SCA, so the responding EMS unit did not have the LUCAS™2 device. In other instances, the LUCAS™2 was deployed but did not operate as long as needed because the battery had not been fully charged.

All these examples pointed to several external elements and agencies involved in determining whether a patient survived an SCA. That is, the LUCAS™2 was not solely responsible for a poor patient outcome, nor could it be assigned all the credit when a patient survived.

RATIONALE FOR A DIFFERENT APPROACH

Many prominent evaluators recognize the importance of accounting for contextual factors beyond the logic model boundaries that affect program success (Chen, 2014; Donaldson, 2007; Pawson, 2013). As Astbury and Hawkins (2019) note, the ways in which evaluators account for context varies as a function of their theoretical orientation. The experimentalist approach to evaluation follows the rationalist position in philosophy and science (Verheij, 2005) and sees context as an unwelcome intruder on experiments designed to reveal some underlying truth about the nature an initiative or action. Here, context is something to be controlled experimentally or partialed out statistically to allow the true 'platonic essence' of

an intervention to be revealed—an approach found to be useful in some, but not all, attempts at generating knowledge in complex systems ([Hawkins, 2016](#)).

The realist approach sees context as a co-conspirator or accessory that is an integral part of the causal process. On the realist account it is the latent, dormant, hidden, or abstract causal mechanisms that do the work, not the interventions we observe in the world ([Pawson & Tilley, 1997](#)). These causal mechanisms are inferred to arise from the relations among components of structures that make up the world. They exist apart from any intervention that may leverage them ([Collier, 1994](#)). Context on this account is something to be harnessed and must be explained, understood, and incorporated in any attempt to ensure that an initiative or action has an effect.

The systems approach is more enamored with context than with either the cold experimental intervention or the affectionate realist mechanism. To a systems evaluator, context is the starting point ([Renger, 2015](#)). It is the main thing that needs to be understood prior to considering the value of any intervention. Here there is often little use for the search for stable and cause-and-effect relations—the focus is on real-time data collection and decision making to improve the efficiency of the system, and sometimes to change it in fundamental ways.

Using a more pragmatic approach, [Rogers \(2008\)](#) suggests adding causal strands as a way of capturing contextual factors influencing program success. In essence, the impact of each contextual factor is accounted for by adding another logic model. Our concern with using this approach was that it did not align with the fundamental characteristics of the cardiac care system. Many agencies and elements (e.g., LUCAS™2) must work together in a coordinated, connected way to improve patient outcomes. Logic models, no matter how many layers they contain, are not designed to capture this connectivity and are often preferred by evaluators because it is human nature to avoid complexity ([Hummelrunner, 2011](#)). Alternatively, [Williams and Hummelrunner \(2010\)](#) published a compilation of systems thinking methods (e.g., social network analysis, causal loops, outcome mapping, etc.) that evaluators might add to their toolkit to augment a program evaluation. Williams and Hummelrunner suggest that many of these research methods might be useful in the program evaluation context.

After reviewing numerous approaches, we decided that a systems theory approach best suited the cardiac care system characteristics: the key that best fitted the lock ([Williams, 2010](#)). In the authors' judgment, the other approaches were an attempt either to adapt methods with which an evaluator was familiar and/or most comfortable or to compensate for a less than ideal methodology.

SYSTEM EVALUATION APPROACH

We chose system evaluation theory (SET) because it is an *evaluation* theory that incorporates systems concepts and guides evaluators on when and how to use them ([Renger, 2015](#)). Further, to the authors' knowledge at the time of the

evaluation there was no other theory specifically designed to evaluate a system as the evaluand.

SET consists of three steps: define the system, evaluate system efficiency, and evaluate effectiveness. Step 1 incorporates various system actor perspectives to define the boundaries, elements, and interrelationships (Renger, Foltysova, Renger, & Booze, 2017). Defining a system is a critical first step because it is necessary to understand what a system looks like and how it is supposed to operate before it can be evaluated. The system structure and functioning form the standard of acceptability against which the evaluation can occur (Green, Kreuter, Deeds, Partridge, & Bartlett, 1980). SET's process for defining the system can be likened to that of building an interlocking jigsaw puzzle. You begin by first building the border (i.e., boundary) and then fill in the pieces (i.e., elements) to solve how they interlock (i.e., interrelationships).

We used the perspectives of state emergency-response leaders to define the cardiac care system boundaries. State leadership was deliberately involved because the research repeatedly shows the need for buy-in at the highest level for both system and evaluation success (Mohan, 2014). We began by first asking leaders to define the goal of the cardiac care system. The agreed-upon ascribed goal was “getting the patient to definitive care in the shortest time possible.” State leadership was then asked to identify cardiac care system elements (e.g., agencies, technology, individuals) that share this goal. As shown in Figure 1, five key agencies were defined: dispatch (law enforcement), the volunteer EMS first responding to the scene, the EMS service with advanced medical-care capabilities (i.e., a paramedic) who then assists in SCA cases, the critical access hospital, and the definitive care facility (i.e., heart hospital). Figure 1 also shows two elements that fell outside the cardiac care system boundaries. This is done purposefully so the reader can better visualize the boundary concept. For the upstream boundary, leadership reasoned

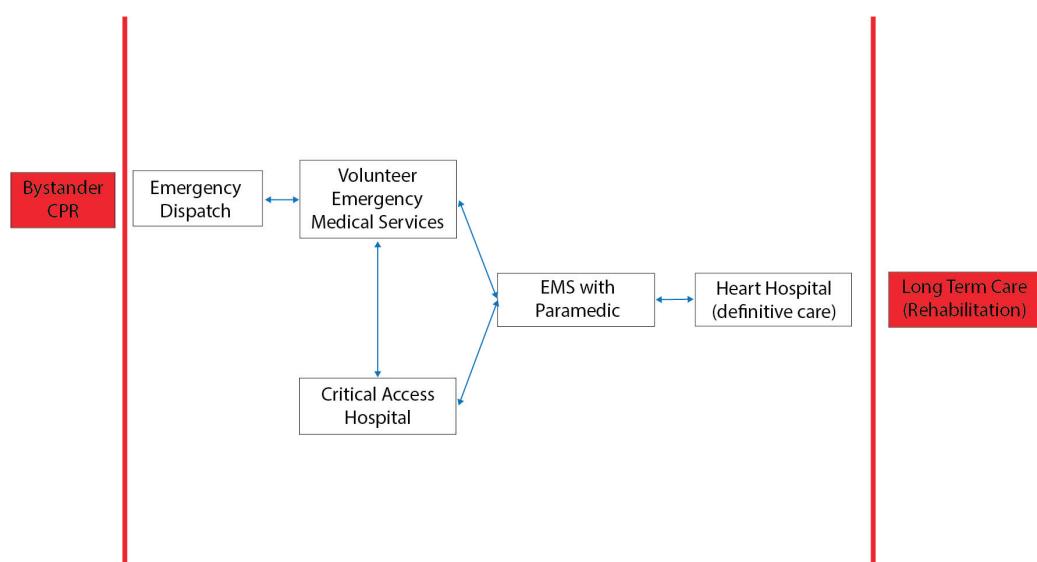


Figure 1. Emergency response boundaries and subsystems

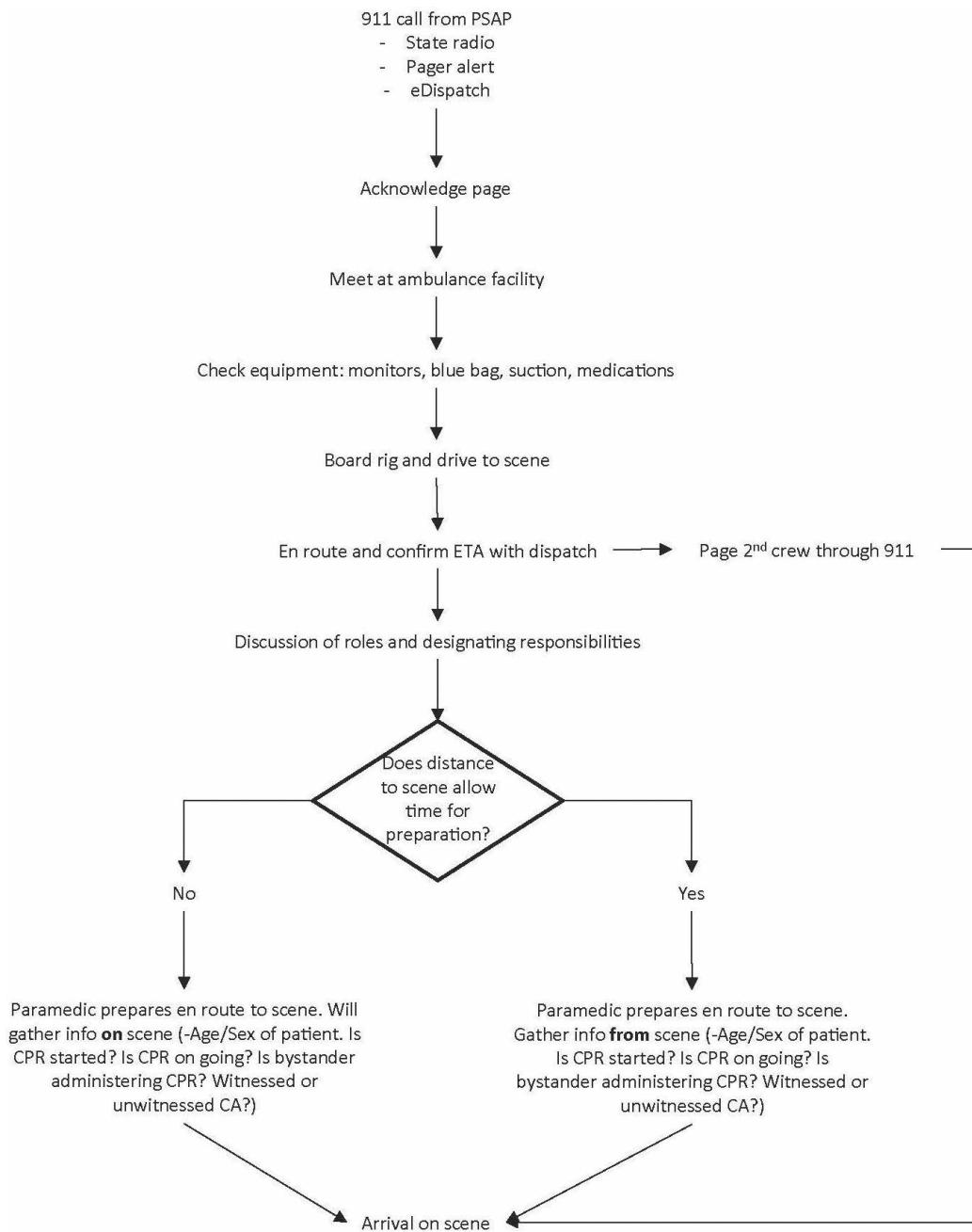
that while CPR-capable bystanders are a key factor affecting patient survival rates, this subsystem preceded the emergency response. With respect to the downstream boundary, the response ended with the patient being treated at a definitive care facility. Rehabilitation and long-term care system elements were considered important for improving a patient's quality of life post-response.

To complete the process of defining the system, we asked state leaders to identify individuals in each subsystem with substantive expertise about the system operations. These operations experts included dispatchers, EMTs, and medical professionals at the CAH and definitive care facility. We then engaged these subject-matter experts in a process flow mapping exercise to detail their standard operating procedures (SOPs) (Renger, McPherson, Kontz-Bartels, & Becker, 2016). A sample SOP is shown in [Figure 2](#).

With the system defined, we proceeded with Step 2 of SET: evaluating system efficiency. We did this by planning, conducting, and evaluating a simulated SCA event (HSEEP, 2013; Manghani, 2011). The simulation exercise mimics an actual event in that people and resources are deployed. The evaluator controls the script, that is, how the exercise unfolds. This level of control gives the evaluator the opportunity to slow down play to better observe or provide players with additional opportunities to learn and/or to speed up play to observe the impact of system stress. After the simulation exercise, we noted where the SOPs were not executed with fidelity. We then systematically applied several systems concepts, including feedback loops, cascading failures, reflex arcs, and system wastes, to better understand why the SOP deviation occurred and to identify potential corrective actions to improve efficiencies.

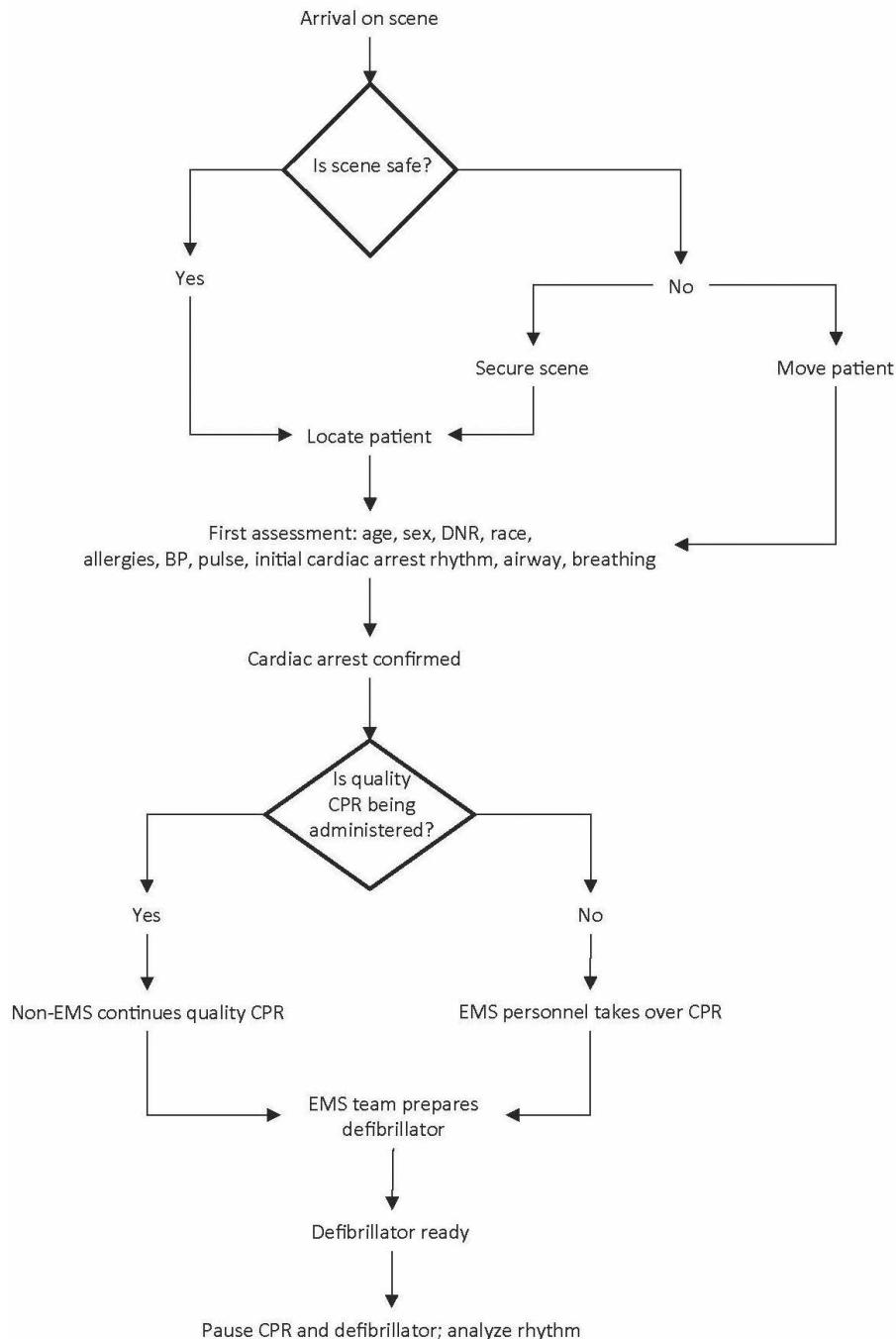
For example, the system concept of waste directed us to inspect whether any SOP steps were being repeated or reworked (Renger, Keogh, et al., 2018). Doing this we found that EMTs and the hospital staff were both calling the same list of health professionals who could assist with medical transport. In addition to wasting time, it was creating confusion among those being called. To resolve this unnecessary redundancy, we brought both parties together in a table-top exercise (HSEEP, 2013). The table-top exercise uses a discussion-based format to simultaneously engage multiple agencies in a low-threat environment. Both agencies were presented with an evolving cardiac event scenario. At each scenario step, agency representatives discussed whether their SOP was able to meet the evolving challenges. As a result, both parties were able to appreciate where their processes overlapped and through discussion resolved the rework issue. The resolution was then reflected in each agency's modified SOP.

We then used the system concept of the reflex arc, a way to improve efficiency by bypassing higher-order functions, to further explore where additional SOP inefficiencies might exist. In practical terms, the reflex arc guided us to look at where SOPs engaged higher organizational levels when the issue could be dealt with more directly at the local level. For example, the evaluation found that important communication was being delayed and decayed because two different dispatching agencies were being used to patch communications between the responding and

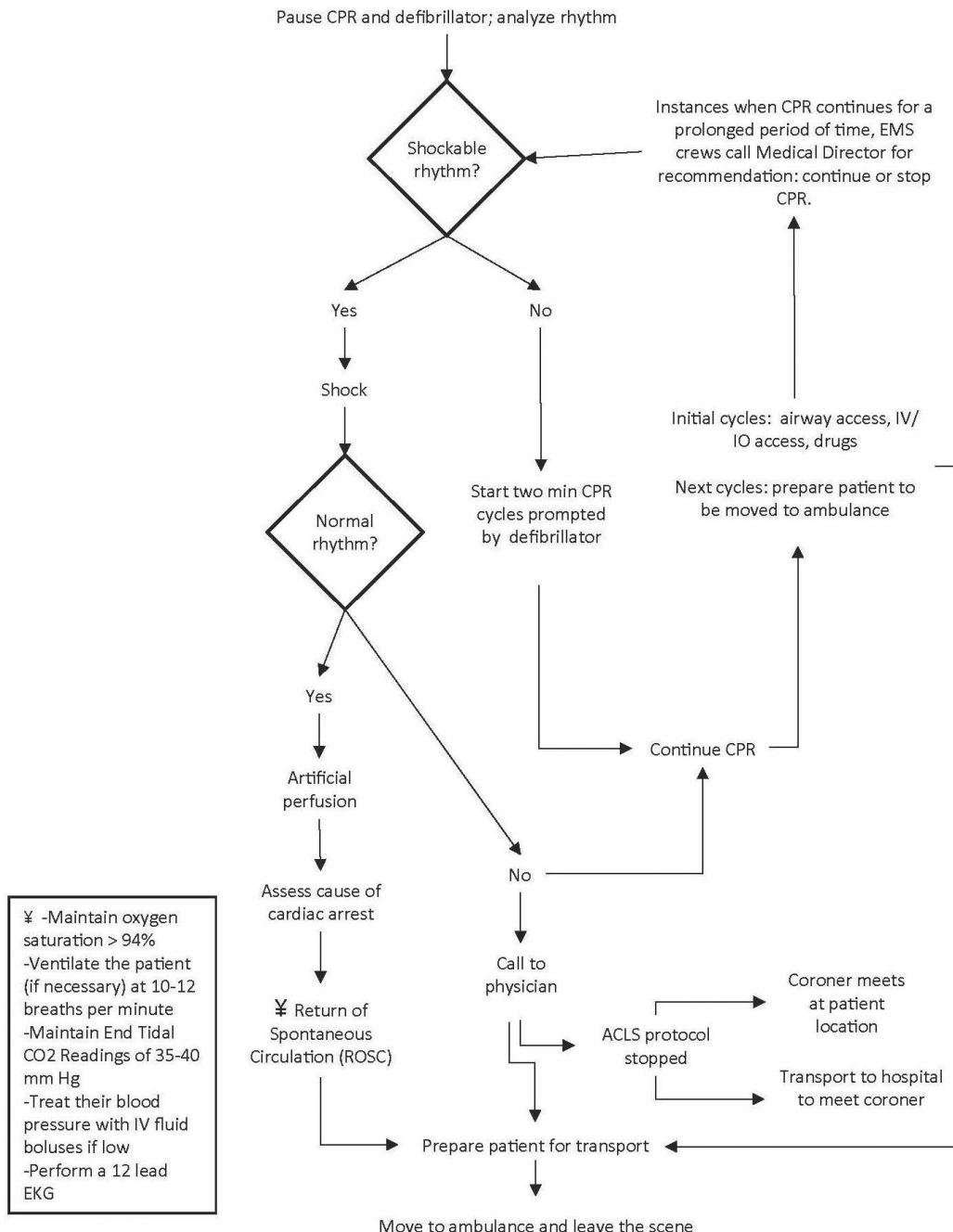
**Figure 2.** Sample SOP

intercepting ambulance (Renger, Harness, Souvannasacd, & Granillo, 2018). The solution was to change the SOPs such that the drivers of the respective ambulances could communicate directly with each other through a dedicated, statewide emergency network.

Feedback loops were another system concept we used to evaluate SOP efficiency. Feedback loops are essential to the evaluating system's interconnectedness (Renger, 2016; Rogers, 2008; Williams & Hummelbrunner, 2010). We found

**Figure 2.** (Continued)

several instances where the system response depended on communication and/or information technology feedback loops. When a feedback loop was identified, it was evaluated using six criteria. First, we evaluated whether the loop was being closed; that is, we ensured that information was delivered to its intended target. Once we confirmed that the feedback loop was closed, we evaluated the feedback quality using five criteria: specificity, timeliness, sufficient frequency, credibility,

**Figure 2.** (Continued)

and relevance (Department of Defense, 2016; Renger, 2016). Not all criteria are germane to every feedback loop. By applying these criteria, we found that the feedback loop between dispatchers and EMTs suffered from a lack of specificity. The chief complaint from EMTs was that dispatchers, who are law-enforcement trained, provided inadequate details about a patient's condition. In response to this feedback specificity issue, dispatchers are now receiving training regarding how to better triage the call to solicit more meaningful information needed

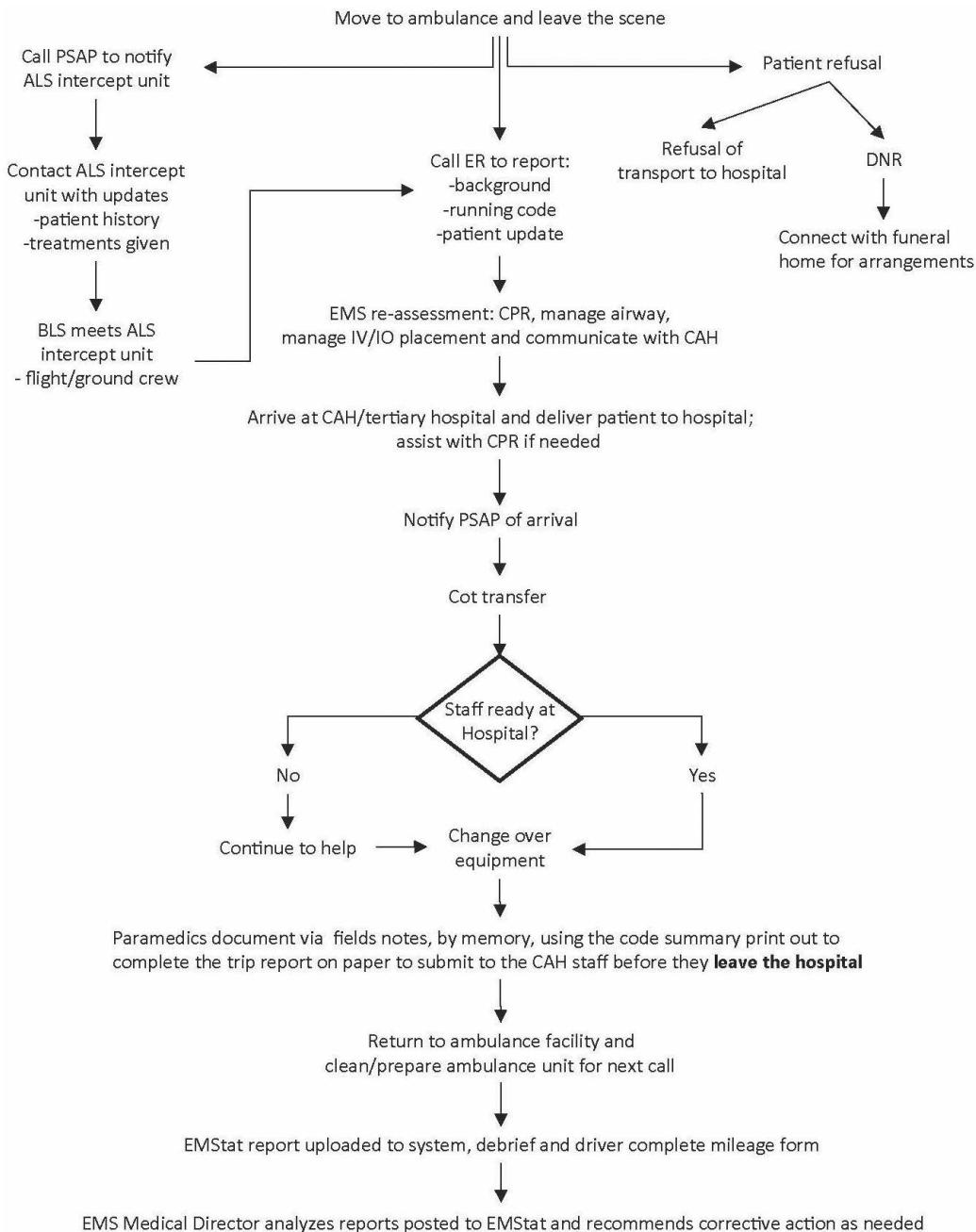


Figure 2. (Continued)

downstream. In evaluating another feedback loop, we learned that EMTs were not providing regular five-minute updates, as per their SOPs, to the receiving health facility. This was an example of information not being provided with sufficient frequency to enable the receiving hospital to appropriately prepare for the incoming patient. This problem was corrected through additional, targeted EMT training.

Finally, we examined SOP deviations through the systems concept lens of cascading failures. This systems concept helped us understand that because system

elements are interconnected, we may need to look upstream from an observed problem to establish its cause. For example, during one of the simulated exercises we learned that the definitive care facility was not beginning preparation for the incoming patient ([Granillo, Renger, McPherson, Dalbey, & Foltysova, 2014](#)). We then followed the SOPs upstream and learned that the definitive care facility was waiting to be contacted by the CAH, which in turn was waiting for the EMS to send an EKG. The EMS hadn't sent the EKG because it didn't have a wifi/data connection. As a result, there was a time-delaying cascading, or domino, effect through the system. The corrective action was to map areas of connectivity in rural areas and change the SOP such that EMS personnel could proceed to the known areas of connectivity as soon as possible to transmit their vital patient data ([Harness et al., 2019](#)).

SET Step 3 is to evaluate system effectiveness. SET seeks to evaluate efficiency before effectiveness—an order that may appear backwards in a program evaluation. Why would we seek to make something more efficient if we are unsure of its effectiveness? The reasoning lies in the different nature of the evaluands. First, a system is pre-existing and generally complex and adaptive, including uncertain interdependencies. We encounter systems; we don't impose them. Second, any systems evaluation has been bounded somewhat artificially for the purposes of evaluation. The effectiveness of the system will in large part be a product of how we have defined the system. A systems evaluation may more usefully determine how well the parts of what has been defined work together to some outcome of interest. In a systems evaluation, a changing problem condition is the outcome—there is no pressing need for attribution to any given intervention or component of the system. A program evaluation, on the other hand, is seeking to warrant the effectiveness of an intervention that has been imposed on a system. It would be unethical and illogical to implement a highly efficient yet ineffective intervention. Systems-thinking concepts of wholeness and interconnectedness are central to evaluating system effectiveness.

DISCUSSION

In the selection of an evaluation approach, it is important to remember that the key must fit the lock ([Williams, 2010](#)). In some cases, where the theory is straightforward, a program logic model is appropriate, and evaluators shouldn't over-engineer or complicate evaluations by applying systems thinking to a linear problem. On the other hand, if an evaluator wants to understand a constellation of interconnecting factors influencing an outcome, then a systems approach, like SET, will likely be better suited to provide meaningful information to help decision makers improve system efficiency and effectiveness ([Astbury & Hawkins, 2019](#)).

Applying systems concepts proved very useful in understanding where to examine system SOPs for interconnected efficiencies. There is a growing body of research devoted to the study of each of the aforementioned system concepts.

SET's unique contribution lies in how it guides evaluators to systematically use system concepts to ensure that system efficiency is evaluated from multiple perspectives. Interconnectedness is inherent in the system concepts of feedback loops, reflex arcs, and cascading failures. However, each system concept addresses a different connectedness aspect. It is necessary to apply all these systems concepts to conduct a comprehensive evaluation that produces the best possible system improvement recommendations.

As the case example illustrates, the process/efficiency evaluation of a program versus a system are very different. SET guides the evaluator in considering the processes within and between system elements. Program evaluation process evaluations are focused on answering whether the intervention is being delivered with fidelity. Evaluation questions about how a program's protocol may interface with other protocols fall outside of this evaluation scope.

With respect to evaluating impact/effectiveness, there is no discernable difference between program and system evaluation outcomes. This is because SET views the attempt to partition or attribute success differentially to parts of the system as meaningless. System elements are interconnected, working together holistically and synergistically. Success is thus measured at a system, not a subsystem, level. Where SET differs is being able to better account for an array of efficiency-related system elements that can influence effectiveness. In this case example, therefore, better patient SCA outcomes were made possible because of a more comprehensive evaluation of how system elements work together to produce the outcome.

Moving from a program evaluation to a systems evaluation approach increases the evaluation scope and cost. The ability of the evaluator to embrace a better-suited approach will be dependent on funding. We were grateful that our funder was flexible in allowing us to change to a better-suited evaluation approach. We were also appreciative that the funder was able and willing to provide the funding necessary to complete the systems evaluation. Other evaluators considering a systems evaluation would benefit from first defining the system so that the scope and cost can be established before proceeding with the evaluation itself.

ACKNOWLEDGMENTS

Dr. Gary Hart passed away before he could see this article into press. I am honored to be a co-author on what turned out to be his last in a very long list of scientific contributions. I will miss his guidance, support, and thirst for knowledge. He was a valuable sounding board for all of my crazy new ideas. Most of all I will miss his friendship. My love and prayers to his family who undoubtedly will miss him more than me. (Ralph Renger)

REFERENCES

- Ackoff, R. L. (1971). Towards a system of systems concepts. *Management Science*, 7(11), 661–671. <https://doi.org/10.1287/mnsc.17.11.661>

- American Evaluation Association (AEA). (2018). *Principles for effective use of systems thinking in evaluation*. Systems in Evaluation TIG. Retrieved from <https://www.systemsinevaluation.com/wp-content/uploads/2018/10/SETIG-Principles-FINAL-DRAFT-2018-9-9.pdf>
- Astbury, B., & Hawkins, A. (2019). *The consolations of evaluation theory*. Presentation given at the Australian Evaluation Society Annual Conference, Sydney, Australia.
- Bohm, K., Rosenqvist, M., Herlitz, J., Hollenberg, J., & Svensson, L. (2007). Survival is similar after standard treatment and chest compression only in out-of-hospital bystander cardiopulmonary resuscitation. *Circulation*, 116, 2908–2912. Retrieved <https://aha.journals.org/doi/full/10.1161/circulationaha.107.710194>
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chen, H. T. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated perspective*. Newbury Park, CA: Sage.
- Collier, A. (1994). *Critical realism: An introduction to Roy Bhaskar's philosophy*. London, England: Verso.
- Department of Defense. (2016). *Relationship between effective performance management and performance feedback*. Retrieved from <https://slideplayer.com/slide/14350612/89/images/6/and+Performance+Feedback.jpg>
- Donaldson, S. I. (2003). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium*. Mahwah, NJ: Lawrence Erlbaum.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York, NY: Psychology Press.
- Eisenberg, M. (2013). *Resuscitate! How your community can improve survival from sudden cardiac arrest* (2nd ed.). Seattle, WA: University of Washington Press.
- Gesellschaft für Internationale Zusammenarbeit (GIZ). (2011). GIZ International Conference on Systemic Approaches in Evaluation, Eschborn, Germany, January 25-26. Retrieved from <https://sites.google.com/a/cpwf.info/l2i/blog/gizconferenceonsystemicapproachestoevaluation>
- Granillo, B., Renger, R., McPherson, M., Dalbey, D., & Foltysova, J. (2014). *Redfield-Aberdeen-Sioux falls cardiac arrest drill after-action report/improvement plan* [Unpublished technical report]. Grant Forks, ND: Center for Rural Health, University of North Dakota.
- Green, L. W., Kreuter, M. W., Deeds, S. G., Partridge, K. B., & Bartlett, E. (1980). *Health education planning: A diagnostic approach*. Palo Alto, CA: Mayfield Publishing.
- Guzy, P. M., Pearce, M. L., & Greenfield, S. (1983). The survival benefit of bystander cardiopulmonary resuscitation is a paramedic served metropolitan area. *American Journal of Public Health*, 73(7), 766–769. <https://doi.org/10.2105/ajph.73.7.766>. Medline:6859360
- Harness, D. S., Sullivan, S., Keegan, J., Taylor, C., Neibauer, T., Jones, I., . . . Renger, J. (2019, September 3). Coordinating a ground-air intercept: Lessons learned. *Journal of Emergency Medical Services*. Retrieved from <https://www.jems.com/articles/2019/08/coordinating-ground-air-intercept.html>

- Hawkins, A. J. (2016). Realist evaluation and randomised controlled trials for testing program theory in complex social systems. *Evaluation*, 22(3), 270–285. <https://doi.org/10.1177/1356389016652744>
- Helmsley. (2019). *The Leona M. and Harry B. Charitable Trust Rural Healthcare Program*. Retrieved from <https://helmsleytrust.org/programs/health-rural-healthcare>
- HSEEP. (2013). *Homeland security and exercise evaluation program*. Department of Homeland Security. Retrieved from https://www.fema.gov/media-library-data/20130726-1914-25045-8890/hseep_apr13_.pdf
- Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation*, 17(4), 395–403. <https://doi.org/10.1177/1356389011421935>
- Manghani, K. (2011). Quality assurance: Importance of systems and standard operating procedures. *Quality*, 2(1), 34–37. <https://doi.org/10.4103/2229-3485.76288>. Medline:21584180
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Meadows, D. H., & Wright, D. (2008). *Thinking in systems: A primer*. White River Junction, VT: Chelsea Green Publishing.
- Meaney, P. A., Bobrow, B. J., Mancini, M. E., Christenson, J., de Caen, A. R., Bhanji, F, . . . Leary, M. (2013). Cardiopulmonary resuscitation quality: Improving cardiac resuscitation outcomes both inside and outside the hospital. A consensus statement from the American Heart Association. *Circulation*, 128(4), 417–435. <https://doi.org/10.1161/cir.0b013e31829d8654>. Medline:23801105
- Mohan, R. (2014). Evaluator advocacy: It is all in a day's work. *American Journal of Evaluation*, 35(3), 397–403. <https://doi.org/10.1177/1098214014532167>
- National Institutes of Health (NIH). (2019). Sudden cardiac arrest. Retrieved from <https://www.nhlbi.nih.gov/health-topics/sudden-cardiac-arrest>
- Pawson, R. (2013). *Science of evaluation: A realist manifesto*. London, England: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London, England: Sage.
- Physio-control. (2016). *The LUCAS™2 training manual*. Retrieved from http://www.lucas-cpr.com/web_training_center/en/pdf2/LUCAS_2_Web_Training_Center_CHECK%20LIST_EN_R05.pdf
- Pima County Health Department. (2018). *Program coordinator job description*. Retrieved from <https://www.governmentjobs.com/careers/pima?department%5B0%5D=Health&sort=PositionTitle%7CAscending>
- Renger, R. (2015). System Evaluation Theory (SET): A practical framework for evaluators to meet the challenges of system evaluation. *Evaluation Journal of Australasia*, 15(4), 16–28. <https://doi.org/10.1177/1035719x1501500403>
- Renger, R. (2016). Illustrating the evaluation of system feedback mechanisms using System Evaluation Theory (SET). *Evaluation Journal of Australasia*, 16(4), 15–21. <https://doi.org/10.1177/1035719x1601600403>
- Renger, R., Atkinson, L., Renger, J., Renger, J., & Hart, G. (2019). The connection between logic models and systems thinking concepts. *Evaluation Journal of Australasia*, 19(2), 79–87. <https://doi.org/10.1177/1035719x19853660>

- Renger, R., Foltysova, J., Ienuso, S., Renger, R., & Booze, W. (2017). Evaluating system cascading failures. *Evaluation Journal of Australasia*, 17(2), 29–36. <https://doi.org/10.1177/1035719x1701700205>
- Renger, R., Foltysova, J., Renger, J., & Booze, W. (2017). Defining systems to evaluate system efficiency and effectiveness. *Evaluation Journal of Australasia*, 17(3), 4–13. <https://doi.org/10.1177/1035719x1701700302>
- Renger, R., Harness, D. S., Souvannasacd, E., & Granillo, B. (2018). The ALS intercept: The need for standard operating procedures (SOPs) and mutual aid agreements (MAAs) to improve patient care and attenuate the impact of billing constraints. *Journal of Emergency Medical Services*. Retrieved from <https://www.jems.com/articles/2018/12/rethinking-als-intercepts-and-mutual-aide-agreements.html>
- Renger, R., Keogh, B., Hawkins, A., Foltysova, J., & Souvannasacd, E. (2018). Reworks: A robust system efficiency measure. *Evaluation Journal of Australasia*, 18(3), 183–191. <https://doi.org/10.1177/1035719x18796611>
- Renger, R., McPherson, M., Kontz-Bartels, T., & Becker, K. L. (2016). Process flow mapping for systems improvement: Lessons learned. *Canadian Journal of Program Evaluation*, 31(1), 109–121. <https://doi.org/10.3138/cjpe.267>
- Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *American Journal of Evaluation*, 23(4), 493–503. <https://doi.org/10.1177/109821400202300409>
- Renger, R., Wood, S., Williamson, S., & Krapp, S. (2011). Systemic evaluation, impact evaluation and logic models. *Evaluation Journal of Australasia*, 11(2), 24–30. <https://doi.org/10.1177/1035719x1101100204>
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), 29–48.
- Sandroni, C., & Geocadin, R. G. (2016). Neurological prognostication after cardiac arrest. *Current Opinion in Critical Care*, 21 (3), 209–214. <https://doi.org/10.1097/mcc.0000000000000202>. Medline:25922894
- Sutton, R. M., Maltese, M. R., Niles, D., French, B., Nishisaki, A., Arbogast, K. B., . . . Nadkarni, V. (2009). Quantitative analysis of chest compression interruptions during in-hospital resuscitation of older children and adolescents. *Resuscitation*, 80(11), 1259–1263. <https://doi.org/10.1016/j.resuscitation.2009.08.009>. Medline:19733427
- Ulrich, W. (2002). Boundary critique. In H. G. Daellenbach & R. L. Flood (Eds.), *The informed student guide to management science*. London, England: Thomson Learning.
- Verheij, B. (2005). Evaluating arguments based on Toulmin's scheme. *Argumentation*, 19(3), 347–371. <https://doi.org/10.1007/s10503-005-4421-z>
- Williams, B. (2010). *Fitting the key to the lock: Matching systems methods to evaluation questions*. Paper presented at the 24th Annual Conference of the American Evaluation Association, San Antonio, TX.
- Williams, B., & Britt, H. (2014). *Systemic thinking for monitoring: Attending to interrelationships, perspectives and boundaries*. Washington, DC: United States Agency for International Development. Retrieved from <https://usaidlearninglab.org/library/attending-interrelationships-perspectives-andboundaries-complexity-aware-monitoring>
- Williams, B., & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner's toolkit*. Stanford, CA: Stanford University Press.

AUTHOR INFORMATION

Ralph Renger taught program and system evaluation at the University of Arizona and the University of North Dakota for 23 years. He is now the president of Just Evaluation Services, LLC. He has published over 50 applied articles in the program and system domains to help evaluators transition from theory to practice so that their evaluations produce more meaningful and useful results.

Jessica Renger is a graduate student in the Evaluation and Applied Research Methods PhD program at Claremont Graduate University. Before going to Claremont, she received a bachelor's degree in neuroscience from the University of Arizona. She has prior experience assisting with health-related evaluation projects in Arizona, North Dakota, and Wyoming. She plans to pursue a career as both an evaluation practitioner and researcher in the future.

Stewart I. Donaldson is Distinguished University Professor, Executive Director of the Claremont Evaluation Center, and Director of The Evaluators' Institute (TEI). He has a large evaluation practice, publishes books and articles, and teaches a variety of courses on evaluation theory and science, and was president of the American Evaluation Association during the International Year of Evaluation in 2015.

Jirina Renger is a graduate of the University of Arizona, College of Public Health. She currently serves as a core faculty member in the DrPH and PhD Public Health programs in the College of Health Sciences at Walden University. Her area of specialization is theory-driven evaluation and program theory reconstruction.

Gary Hart was an expert in rural health research with an emphasis in health workforce and program evaluation. He was the director of one of the nation's, largest, oldest, and most prestigious rural centers.

Andrew Hawkins is a partner with ARTD Consultants, one of Australia's most established specialist evaluation firms. He has a keen interest in translating evaluation theory into practice to generate evidence and insight for decision makers.