



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## A conceptual framework for evaluating cooking systems

**Citation for published version:**

Bond, TC, L'Orange, C, Medwell, PR, Sizoomu, G, Abdelnour, S, Brinkmann, V, Lloyd, P & Pemberton-Pigott, C 2022, 'A conceptual framework for evaluating cooking systems', *Environmental Research Letters*, vol. 17, no. 3, 031002. <https://doi.org/10.1088/1748-9326/ac5418>

**Digital Object Identifier (DOI):**

[10.1088/1748-9326/ac5418](https://doi.org/10.1088/1748-9326/ac5418)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Environmental Research Letters

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



PERSPECTIVE • OPEN ACCESS

## A conceptual framework for evaluating cooking systems

To cite this article: Tami C Bond *et al* 2022 *Environ. Res. Lett.* **17** 031002

View the [article online](#) for updates and enhancements.

### You may also like

- [Effect of V/Fe composite interlayer on the microstructure and mechanical properties of electron beam welded joints of Ta-10W and GH3128](#)

Yingnan Li, Dong Li and Zhengyi Cao

- [Non-parametric inference of impurity transport coefficients in the ASDEX Upgrade tokamak](#)

Takashi Nishizawa, Ralph Dux, Rachael M McDermott *et al.*

- [Determination of the responsivity of Predictable Quantum Efficient Detector over a wide spectral range based on a 3D model of charge carrier recombination losses.](#)

Trinh Tran, Geiland Porrovecchio, Marek Smid *et al.*

ENVIRONMENTAL RESEARCH  
LETTERS

## PERSPECTIVE

## A conceptual framework for evaluating cooking systems

## OPEN ACCESS

RECEIVED  
6 April 2021REVISED  
18 January 2022ACCEPTED FOR PUBLICATION  
9 February 2022PUBLISHED  
7 March 2022

Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



Tami C Bond<sup>1,\*</sup> , Christian L'Orange<sup>1</sup> , Paul R Medwell<sup>2</sup> , George Sizoomu<sup>3</sup>, Samer Abdelnour<sup>4</sup> ,  
Verena Brinkmann<sup>5</sup>, Philip Lloyd<sup>6,9</sup> and Crispin Pemberton-Pigott<sup>7,8</sup> 

<sup>1</sup> Mechanical Engineering, Colorado State University, Fort Collins, CO 80521, United States of America

<sup>2</sup> Mechanical Engineering, University of Adelaide, Adelaide, SA 5005, Australia

<sup>3</sup> Racell Uganda Ltd, Kampala, Uganda

<sup>4</sup> Business School, University of Edinburgh, Edinburgh EH8 9JS, United Kingdom

<sup>5</sup> Energising Development, Deutsche Gesellschaft für Internationale Zusammenarbeit, Bonn 53113, Germany

<sup>6</sup> Energy Institute, Cape Peninsula University of Technology, Cape Town 7530, South Africa

<sup>7</sup> College of Engineering, China Agricultural University, Beijing 100083, People's Republic of China

<sup>8</sup> New Dawn Engineering, Manzini MZ 200, Eswatini

<sup>9</sup> Philip Lloyd predeceased the final text of this document and permission to publish was granted by his heirs.

\* Author to whom any correspondence should be addressed.

E-mail: [Tami.Bond@colostate.edu](mailto:Tami.Bond@colostate.edu)

**Keywords:** cooking stoves, household energy, performance metrics, impact evaluation, complex systems

Supplementary material for this article is available [online](#)

## 1. Introduction

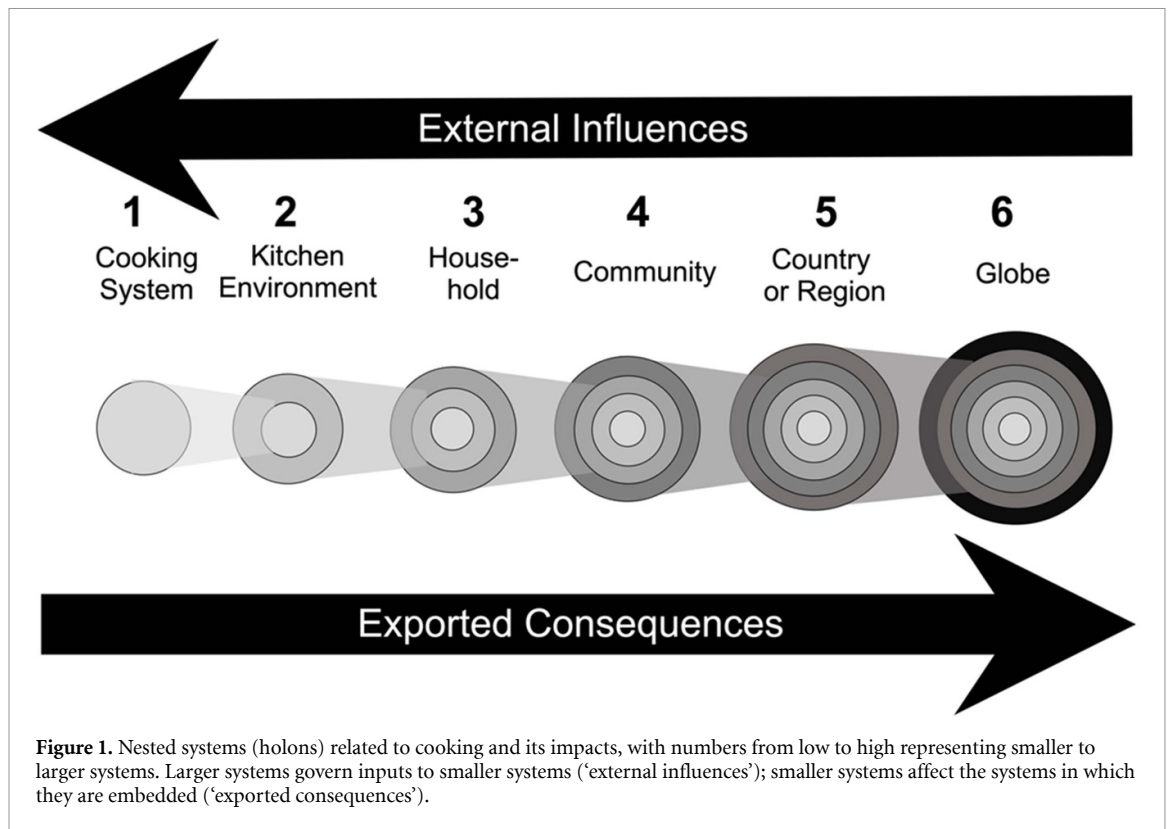
Providing household energy of adequate quality and quantity is a major global challenge, as reflected in discussions on energy access, gender impacts, and health (Bouzarovski and Petrova 2015). The United Nations Sustainable Development Goal 7 expresses a need for 'affordable, reliable, sustainable and modern energy for all' as well as improved infrastructure and technology for delivery (United Nations 2021). Cooking is a basic human need, yet around 2.6 billion people worldwide still rely on solid fuels, such as wood, dung, agricultural residues, coal, and charcoal (IEA, IRENA, UNSD, World Bank 2021, Stoner *et al* 2021). These fuels and devices may degrade natural resources, expose users to high levels of emissions that affect health and climate (Naeher *et al* 2007, Lacey *et al* 2017, Kodros *et al* 2018), and consume more money, time and effort than more effective alternatives (Sagar 2005, Jeuland and Pattanayak 2012).

It is tempting to believe that clean and efficient cooking can be easily achieved by changing stoves, fuels and kitchens. One challenge is confirming that interventions meet targets in terms of service level, cleanliness, fuel consumption, user preferences, and gender impacts (Barnes *et al* 1994, Ezzati and Kammen 2002, Brooks *et al* 2016). These confirmations come about through performance testing that gives information about fuels or stoves, or through project evaluation, which assesses the outcome of interventions. However, cooking systems function in complex situations. The task of cooking combines one or more cooking devices, their

associated fuels, foods prepared within specific vessels, and operational sequences. These elements are immersed in and affected by the surrounding kitchens or cooking places, types of housing, household dynamics, community structures, and cultures and cooking practices. Although the cooking system has complex relationships with its surroundings, evaluation often focuses on discrete systems.

Procedures for evaluating household energy interventions are widely used (Visser 2005, Makonese *et al* 2012, Arora *et al* 2014, ISO TC 285 2018). There has also been work to illuminate how technical performance or programmatic outcomes relate to social or environmental goals (Glasgow *et al* 1999, Quinn *et al* 2018, World Health Organization 2018, 2019). However, household energy systems are complex (Malhotra *et al* 2004, Guruswamy 2015, Jagadish and Dwivedi 2018), and a structured approach has not yet accounted for this complexity. Hence, there is little guidance for identifying and choosing evaluation procedures that fit specific contexts and purposes, and determining when claims of performance outcomes can be made with confidence. Thus, even widely accepted testing or evaluation procedures may be selected without full attention to their contexts or limitations (Abdelnour *et al* 2020).

Conceptual frameworks have appeared in several fields (Gartner 1985, Hobbs and Norton 1996, Seuring and Müller 2008) to describe systems that interact with and influence each other. This document presents such a conceptual framework for evaluating cooking systems. Its intention is not to answer every stakeholder's question, nor to comment on any



particular evaluation protocol, but rather to present a systematic framework that identifies key concepts for assessing performance and drawing implications. This presentation begins by delineating nested systems that affect and are affected by cooking systems. We use this context to describe the relationship between desired impact and programmatic action. A series of 12 questions is provided to guide stakeholders through the tasks of identifying appropriate performance metrics and targets, determining when a particular testing approach is suitable for a given situation, and interpreting measured performance data. Finally, we reflect on the implications for change and its evaluation in complex systems. For readers who prefer practical guidance rather than conceptual discussion, we provide specific descriptions of the nested systems (appendix A available online at [stacks.iop.org/ERL/17/031002/mmedia](https://stacks.iop.org/ERL/17/031002/mmedia)), a flow chart to outline supportable conclusions (appendix B), a case study that answers the 12 questions in a common situation (appendix C), and a more complex example that builds on the case study situation (appendix D).

## 2. Evaluation in nested systems

Figure 1 shows how systems directly related to cooking are nested within larger systems. Such communicating, nested systems are sometimes called 'holons' (Koestler 1970, Edwards 2005, Bland and Bell 2007). In the framework presented here, each system is delineated by a perceived sphere of influence. The cooking

system (system 1) includes a cooking device, operating controls, a transformation mechanism for a fuel or energy source, and a place for a cooking vessel. This device is created by a manufacturer, who attempts to deliver design-related performance. The cooking system operates within a kitchen environment (system 2), where a user operates the controls, adds fuel, and employs a cooking vessel.

Performance in the kitchen is within the sphere of the stove user. The kitchen environment must serve the needs and work within the constraints of the household (system 3); the household is part of a community with culture, local markets, and political leadership (system 4). Local or national authorities address environmental challenges and health within a region (system 5) and international or multinational organizations attempt to address concerns that affect the entire globe (system 6). Although this work focuses on evaluation, these definitions would also aid in defining the distribution of costs and benefits.

### 2.1. External influences

Even when a party is responsible for a system, that party does not have full control over its performance. Larger systems influence smaller systems by constraining factors that lie outside, but affect the performance of the smaller embedded system. For example, a manufacturer cannot control the quality of fuel used in a stove or the user's operational choices, and occupants cannot prevent the infiltration of polluted air into the household.

**Box 1. Terms describing relationships among nested systems.**

See table 1 for examples of systems and metrics. *Correlative relationship.* The relationship between a proxy metric and the desired outcome.

*Exported consequence.* An effect that a smaller system has on a larger system in which it is embedded.

*Desired outcome.* The aspirational outcome that stakeholders wish to achieve in the system of desired impact, which may be very difficult to measure.

*External influence.* An effect that a larger system has on a smaller, embedded system.

*Mechanistic relationship.* The predictable connection between the performance metric and the proxy metric.

*Neighbouring contributions.* A change in the system of desired impact, including in the proxy metric, caused by systems at the same level as the system of action

*Performance metric.* A measurable quantity characterizing the behavior of the system of action.

*Proxy metric.* A measurable quantity in the system of desired impact, which is related to the desired outcome.

*System of action.* The system that the stakeholder can change.

*System of desired impact.* The system that the stakeholder wishes to change.

## 2.2. Exported consequences

Each large system contains and is affected by many smaller systems, just as a body is affected by its internal organs. The party responsible for the major system often has little control over the sub-systems. A cook chooses how to operate a stove, but usually cannot alter the stove's basic design or its performance, although it affects her cooking. A region's environmental ministry may not have authority to regulate domestic cooking, but is still responsible for the airshed containing the pollutants.

## 2.3. Terminology

Motivations for changing small systems are often the consequences they export to larger systems. To describe the interplay among systems, we introduce new terms that are summarized in box 1 and illustrated in figure 2. Aspirations often relate to a larger system ('system of desired impact').

Whether the stakeholder is the stove user who seeks to reduce illness or household expenses, a

public-health official hoping to improve living conditions in a community, a donor program intending to stimulate markets for better products, or a vendor promoting a cooking fuel, each stakeholder hopes that changes will lead to a desired outcome. However, many of these desired outcomes are difficult or impossible to measure directly. Changing the larger, more complex system is usually impractical or impossible, so actors often focus on changing a small system ('system of action') whose behavior can be characterized with one or more performance metrics.

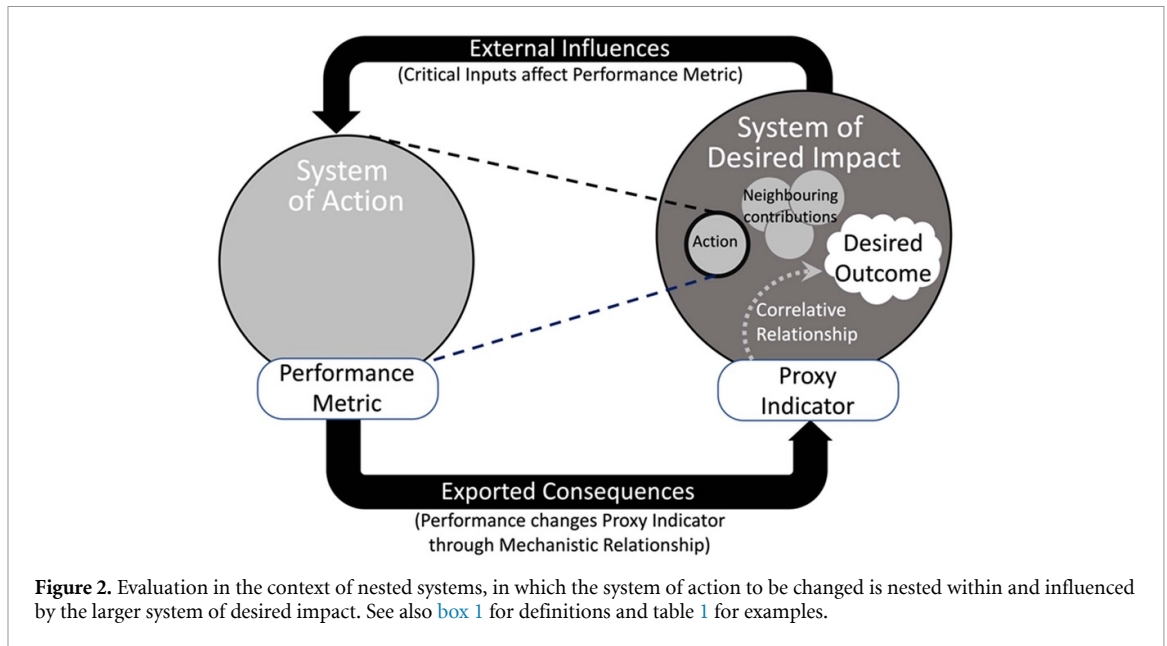
Two connections are needed to ensure that the change in the system of action could improve the system of desired impact. First, if the desired outcome is not measurable, the system of desired impact must contain a quantity that is measurable and that correlates with the desired outcome. We refer to this measurable quantity as the proxy indicator, and its relationship with the desired outcome as the correlative relationship. There can be no confidence that change has actually occurred without an observable indicator of change in the system of desired impact, even if the proxy indicator is not measured in every project. Second, the performance indicator must have a mechanistic relationship with the proxy indicator. That is, the connection must be objectively observable and quantifiable; a hypothetical relationship is insufficient. This connection links the smaller system with the larger one. This assumption, that changes in the system of action lead to changes in the system of desired impact, is a fundamental theory of change that often goes unstated. Further, when these connections are obscured and their uncertainties are not acknowledged, expectations about changes in the system of desired impact may be unrealistic. The system of desired impact also contains other, smaller systems at the same level as the system of action, that may also affect the proxy indicator. We term these systems neighbouring contributions; they may or may not be affected by changes imposed on the system of action.

## 3. Guidance for evaluation

Evaluation procedures can quantify baseline situations, assess whether the desired outcome is likely to be achieved before the activities are carried out, or determine what was achieved after the interventions. Evaluation can be confounded by external influences and by uncertainty in the evaluation process. The 12 questions listed in box 2 are provided to aid in identifying the evaluator's goals, determining how to assess the appropriateness of evaluation mechanisms, and interpreting the results of evaluation procedures.

### 3.1. Identifying systems (Q1–Q2)

The system of desired impact and the system of action depend on the stakeholder seeking the information and the desired outcome of the planned program.



### Box 2. Questions to clarify the purpose and interpretation of evaluation.

- Q1. What is the system of desired impact?  
Q1a. Which stakeholders have an interest in altering that system?
- Q2. What is the system of action?  
Q2a. Which stakeholders alter the system of action and conduct the evaluation?
- Q3. What is the desired outcome?  
Q3a. Which stakeholders have defined the desired outcome?
- Q4. What is the proxy metric?  
Q4a. How is the proxy metric related to the desired outcome (correlative relationship)?  
Q4b. What uncertainty is associated with that relationship?
- Q5. What is the performance metric?  
Q5a. How is the performance metric related to the proxy metric (mechanistic relationship)?  
Q5b. What uncertainty is associated with that relationship?
- Q6. What are the critical inputs?  
Q6a. How do the critical inputs vary within the system of desired impact?
- Q7. What is the target value of the performance metric?  
Q7a. How has the target value been chosen?
- Q8. What level of confidence in conclusions is desired?
- Q9. What is the measured value of the performance metric and its uncertainty?  
Q9a. Which evaluation protocol is fit for the purpose of assessing the value of the performance metric?  
Q9b. What is the value of the performance metric, measured with the chosen protocol?  
Q9c. What is the uncertainty in the performance metric due to random errors, systematic errors, inherent variability, and critical inputs?
- Q10. Was the performance metric target value achieved within the system of action with the required level of confidence?
- Q11. What are the predicted changes in the proxy metric and the desired outcome?  
Q11a. What factors outside the system of action affect the proxy metric and the desired outcome (neighbouring contributions, critical inputs)?  
Q11b. What total uncertainty is associated with the proxy metric?  
Q11c. What total uncertainty is associated with the desired outcome?
- Q12. Is an observable change expected in the system of desired impact?

The system of action is the largest system directly altered by the program. The system of desired impact is typically much larger and more complex than the

system of action. [Table 1](#) lists some possible systems of desired impact and of action relevant to different stakeholders.



**Table 1.** System, metric and input delineation for sample information consumers. The proxy metrics listed here correspond only to the stated desired outcome. This table is intended to be illustrative, not comprehensive; for example, PM is not the only pollutant that affects health.

Information consumer	System of desired impact (SoDI); desired outcome (DO); proxy metric	System of action (SoA); performance metric	Critical inputs (CI); neighbouring contributions (NC)
Stove user	SoDI: household DO: reduce expenditure on fuel Metric: fuel consumption	SoA: cooking system Metric: cost per meal cooked	CI: cooking task, e.g. boiling water <sup>a</sup> ; fuel quality and size NC: non-cooking tasks <sup>b</sup> , e.g. space heating
Public health program	SoDI: region DO: reduce adverse health impacts Metric: exposure of individuals to PM	SoA: kitchen environment Metric: indoor concentration of PM	CI: cooking task; user operation; choice and longevity of low-emitting versus polluting devices; fuel quality and size, kitchen design NC: non-cooking sources of PM
Stove designer	SoDI: region (market) DO: sell devices that meet standards Metric: approval of standard-setting organization	SoA: cooking system Metric: meeting benchmarks such as maximum emission per task and safety measures	CI: inputs specified by standard procedures NC: competing vendors
Carbon emission certificate vendor	SoDI: globe DO: reduce climate impact Metric: greenhouse gas concentration	SoA: household environment Metric: fuel consumption	CI: cooking and non-cooking tasks NC: changes in other sectors such as vehicles and industrial activity

<sup>a</sup> A cooking task is a function that the intervention is intended to replace, and that affects the desired outcome.

<sup>b</sup> A non-cooking task is a function that the intervention is not intended to replace, but that also affects the desired outcome.

### 3.2. Identifying desired outcomes, metrics, and their relationships (Q3–Q5)

Table 1 illustrates a desired outcome, proxy metric, and performance metric for some possible situations. It is important to differentiate between the three outcomes: the desired outcome (Q3) which occurs in the system of desired impact and may not be measurable; the performance metric (Q5) that characterizes the system of action; and the proxy metric (Q4) that is an intermediate step between the performance metric and the desired outcome.

### 3.3. Identifying critical inputs and their influence (Q6)

Because the system of action operates inside the system of desired impact, it must be evaluated under conditions similar to those occurring throughout the system of desired impact. Specifically, inputs that affect the performance metric—‘critical inputs’—must be comparable between the test used for evaluation and those encountered in the system of desired impact. Otherwise, the performance metric fails to represent the performance of the system of action as it is commonly used.

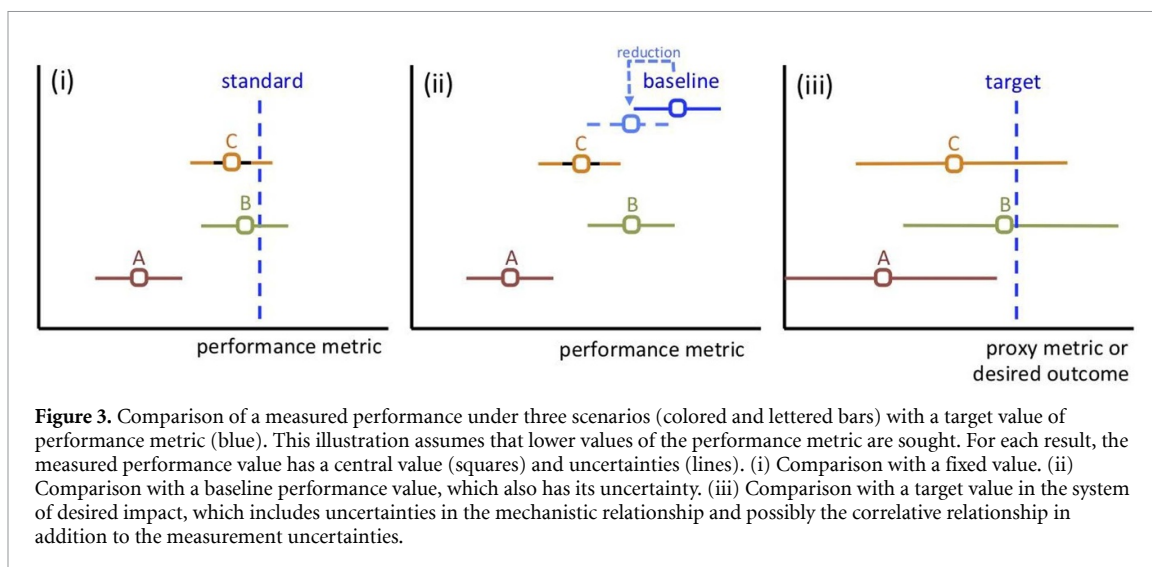
The influence of critical inputs can be determined through sensitivity studies or by measuring the performance metric in naturally-occurring conditions throughout a larger system. This task is best suited to developers of a measurement protocol in consultation with implementers and other stakeholders, and is not well suited to individual programs or evaluators.

The answers to question 6 capture what has been called a ‘controlled versus uncontrolled’ dichotomy. A designer might test a device to ensure its performance under well-controlled conditions, ignoring external influences. For public health evaluations, however, the entire community may be the system of desired impact, adding factors that influence stove performance. This difference has also been called ‘laboratory versus field testing’ (Medina *et al* 2017), although it is caused by variation in critical inputs and imposed controls rather than the physical location of testing.

### 3.4. Identifying performance targets and confidence (Q7–Q8)

Once the performance metric has been identified, the evaluator selects a target value for each metric (Q7) that indicates the success of the project. The statement that the system has achieved the desired level of change must be given with an expressed level of confidence (Q8). There are at least three ways to choose target values, which are illustrated in figure 3 and discussed below. Appendix D contains an example of each method for the case study situation in appendix C.

- (i) Select a statutory or broadly-accepted value such as a performance standard or an environmental standard. These standards might be externally imposed by a government, industry or funding agency.



**Figure 3.** Comparison of a measured performance under three scenarios (colored and lettered bars) with a target value of performance metric (blue). This illustration assumes that lower values of the performance metric are sought. For each result, the measured performance value has a central value (squares) and uncertainties (lines). (i) Comparison with a fixed value. (ii) Comparison with a baseline performance value, which also has its uncertainty. (iii) Comparison with a target value in the system of desired impact, which includes uncertainties in the mechanistic relationship and possibly the correlative relationship in addition to the measurement uncertainties.

- (ii) Compare final performance with a previous result or baseline, for example ‘consumes 25% less fuel than the traditional stove’ or ‘50% less time spent preparing fuel.’ This option requires capturing metric values before and after the intervention. Baselines can be determined through measurement or estimation, but all entail some uncertainty.
- (iii) Relate the performance metric to a desired value of proxy metric. A target proxy metric value is assigned and the equivalent performance metric value is determined by working backward through the mechanistic relationship using equations, statistical relationships or computer simulations. A target desired outcome may also be chosen directly, which would require working backward through the correlative relationship between the desired outcome and the proxy metric. These additional steps introduce greater uncertainty.

Figure 3 illustrates comparisons of the performance metric with central values (squares) and the reported confidence interval (bars) for three hypothetical cooking systems. System A meets the performance metric target value with high confidence, as all likely values meet the target. System B has questionable performance; even though the central value meets the target, many of the likely values do not. The performance of system C is better, yet some of the likely values do not meet the target. However, if the accuracy of the measurement method could be improved (black lines in (i) or (ii)), then there could be greater confidence that system C met the performance target. If the uncertainty in the measurement procedure is large enough, it may not be possible to make a claim of success with the appropriate confidence. In that case, a new procedure should be sought, the apparatus improved, or the target may need revision.

When the target is a proxy metric or desired outcome rather than a performance metric (figure 3(iii)), additional uncertainties can be large. Consequently, there is a lower likelihood of detecting a small change with high confidence.

### 3.5. Measuring the performance metric (Q9)

Once the performance metric, target, and required confidence are determined, a measurement protocol can be sought. Several requirements make a protocol *Fit for Purpose* (De Bievre 2010).

#### 3.5.1. Validity and relevance

The protocol should be valid from a theoretical standpoint, meaning that it must have neither misconceptions in the measurements nor with its interpretation. The protocol should measure a performance metric that is suitably connected to the proxy metric. Relevance indicators may be used to assess whether the protocol is conceptually (internally) consistent, as well as consistent from the perspective of common sense (externally) (Gaskell and Bauer 2000). The validity and relevance of a protocol and a metric usually cannot be assessed by a protocol user; users should look to previous review processes.

#### 3.5.2. Situational appropriateness or contextuality

The protocol should be suited to the skill of the operator and it should be possible to conduct the measurement in the required environment. The instructions given should be appropriate and unambiguous, so that all operators interpret them the same way. The time required to complete the protocol should be justified by the amount of information gained.

#### 3.5.3. Transparency and traceability

The individual executing the protocol must document all choices made so that tests can be replicated. Documentation should include operation of the



measurement, including instrumentation specifications. The documentation should also record choices of critical inputs used during the measurement (Q6), so that a reviewer can assess relevance to the system of desired impact. Confidence indicators may be used to demonstrate that the protocol, its associated measurements, and interpretation of results represent reality (Gaskell and Bauer 2000).

#### 3.5.4. Reporting experimental error

The protocol must report variability inherent in measurements, which is caused by instrumental uncertainty and differences in inputs and operator practice. These errors can be quantified and reduced by examining repeatability and reproducibility. A repeatable measurement produces similar results under the same conditions; a reproducible measurement produces similar results when conducted in different locations or at different times.

#### 3.5.5. Quantifying variability due to critical inputs

The protocol should identify the range of critical inputs expected throughout the system of desired impact, and the reported performance metric should quantify variability caused by that range of critical inputs. For example, if a political district is the system of desired influence, and critical inputs are fuel type and size, the protocol should report variability that would occur with all fuel types and sizes used within the district. This requirement is similar to the idea of external validity (Steckler and McLeroy 2008), but not as stringent; the performance metric must represent the system of action only where it is to be deployed.

#### 3.5.6. Quantifying natural variability in the tested system

The system of action (for example, the cooking system or kitchen environment) often has inherent variability in performance, which can be eliminated only through artificial, strict control. An example is operator practice in feeding solid fuel. Whether variation is attributed to inherent variability or to critical inputs is arguable, but the most important goal is counting all variability under one of the groups.

The criteria listed above are each separately applied. A protocol may be transparent, but not valid to measure the desired performance metric. A protocol may be valid and transparent, but too complex to conduct in a field setting, and thus situationally inappropriate. The protocol may be valid, transparent, and appropriate for the situation, but may not quantify variability sufficiently. Further, including too many diagnostics in a single protocol could cause it to be expensive, cumbersome and limit its situational appropriateness.

There are formal methods of combining variability or uncertainty, whether introduced by experimental error, critical inputs, and natural variability

(Morgan *et al* 1990, Cullen and Frey 1999). The combined uncertainty is to be used in answering the questions that follow.

### 3.6. Interpreting evaluation results for the system of action (Q10)

If the previous choices are clearly answered and well conducted, the user is in a good position to answer Question 10: Has the system of action met the target performance metric with the required degree of confidence?

An affirmative answer to this question means that the system has achieved the target performance metric value. Uncertainties in the measured values, and large variability caused by critical inputs, can prevent the evaluator from being able to answer in the affirmative. Additional uncertainty is incurred when the target is set through a connection to the proxy metric. Methods to reduce uncertainty include (a) increasing robust performance of the system for the full range of critical inputs; (b) decreasing uncertainty inherent in the measurement protocol, but not by reducing the range of critical inputs; and (c) setting a target in the system of action rather than a larger system.

### 3.7. Interpreting evaluation results applicable to the system of desired impact (Q11, Q12)

The performance metric (in the system of action) must have a mechanistic relationship with the proxy metric (in the system of desired impact), and the proxy metric and the desired outcome must have a correlative relationship. However, the proxy metric might also respond to other systems nested in the system of desired impact (neighbouring contributions), and it might have external influences from larger systems. This possibility is even more likely for the desired outcome, as the correlative relationship to the proxy metric introduces yet more variability. Justifiable statements associated with a sequence of findings are summarized here and shown in a flow chart in appendix B.

- If the performance metric has not been tested with the full range of critical inputs, then no conclusions can be made about changes in the system of desired impact.
- If the performance metric represents the full range of critical inputs, and meets the target metric with confidence, then the system of action has achieved the stated performance goal, but an observable change in the system of desired impact is not guaranteed.
- Relative variability in the proxy metric is greater than that in the performance metric (figure 3(iii), Q11b) due to the mechanistic relationship and other influences. Projected uncertainty is needed to determine whether the change is detectable in the system of desired impact.

- Relative variability in the desired outcome is greater than that of the proxy metric (Q11c) due to the correlative relationship and other influences. Assessment of change in the desired outcome is advisable only when uncertainties have been accounted for, and a detectable change is still expected.

It may not be possible to measure a change in the system of desired impact, even if the system of action meets its performance target. This is a common conundrum. The larger system of desired impact might have provided the motivation for change, but realistic programs produce changes that may be observable only at the level of the smaller system affecting it.

#### 4. Summary and outlook

This document proposes a framework for appropriate evaluation of cooking systems. We guide evaluators to develop quantitative theories of change that connects the performance of a small system of action to proxies and desired outcomes in a larger system, outlining features of evaluation protocols that render them fit for the purpose of gauging performance. Assessing uncertainty within the system of action, and of influences external to that system, is key to choosing appropriate protocols and to drawing inferences about performance and impacts. When the questions outlined are answered carefully, evaluators, policymakers and decisionmakers can gain confidence in the reported results and their interpretation.

Uncertainties can confound the detection of change, especially in the system of desired impact. This confounding of outcomes limits the ability to conduct well-accepted types of assessment, including randomized controlled trials. It has long been acknowledged that flexibility in design is needed for interventions in complex systems (Shiell *et al* 2008). The framework presented here demonstrates that assessing outcomes in complex, nested systems also necessitates adaptability. Positive change in complex systems with diverse influences requires systematic methods of recognizing incremental but meaningful contributions, as well as acknowledgement and acceptance of situations when detection of significant change is not feasible.

#### Data availability statement

No new data were created or analysed in this study.

#### Acknowledgments

We are grateful to six reviewers whose perspectives improved the clarity of this presentation. T C B

acknowledges funding for related work under Assistance Agreement No. EPA RD-83542301 awarded by the U.S. Environmental Protection Agency. The document has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the agency.

#### ORCID iDs

Tami C Bond  <https://orcid.org/0000-0001-5968-8928>

Christian L'Orange  <https://orcid.org/0000-0003-1905-8694>

Paul R Medwell  <https://orcid.org/0000-0002-2216-3033>

Samer Abdelnour  <https://orcid.org/0000-0001-5637-8692>

Crispin Pemberton-Pigott  <https://orcid.org/0000-0002-1950-4642>

#### References

- Abdelnour S, Pemberton-Pigott C and Deichmann D 2020 Clean cooking interventions: towards user-centred contexts of use design *Energy Res. Soc. Sci.* **70** 101758
- Arora P, Das P, Jain S and Kishore V V N 2014 A laboratory based comparative study of Indian biomass cookstove testing protocol and water boiling test *Energy Sustain. Dev.* **21** 81–88
- Barnes D F, Openshaw K, Smith K R and van der Plas R 1994 *What Makes People Cook with Improved Biomass Stoves?* (Washington, DC: The World Bank)
- Bland W L and Bell M M 2007 A holon approach to agroecology *Int. J. Agric. Sustain.* **5** 280–94
- Bouzarovski S and Petrova S 2015 A global perspective on domestic energy deprivation: overcoming the energy poverty–fuel poverty binary *Energy Res. Soc. Sci.* **10** 31–40
- Brooks N, Bhojvaid V, Jeuland M A, Lewis J J, Patange O and Pattanayak S K 2016 How much do alternative cookstoves reduce biomass fuel use? Evidence from North India *Resour. Energy Econ.* **43** 153–71
- Cullen A C and Frey H C 1999 *Probabilistic Techniques in Exposure Assessment* (New York: Plenum)
- De Bievre P 2010 'Fitness-for-intended-use' is an important concept in measurement *Accredit. Qual. Assur.* **15** 545–6
- Edwards M G 2005 The integral holon: a holonomic approach to organisational change and transformation *J. Organ. Change Manage.* **18** 269–88
- Ezzati M and Kammen D M 2002 Household energy, indoor air pollution, and health in developing countries: knowledge base for effective interventions *Annu. Rev. Environ. Resour.* **27** 233–70
- Gartner W B 1985 A conceptual framework for describing the phenomenon of new venture creation *Acad. Manage. Rev.* **10** 696–706
- Gaskell G and Bauer M W 2000 Towards public accountability: beyond sampling, reliability and validity *Qualitative Researching with Text, Image and Sound* (London: Sage Publishing Ltd) pp 336–50
- Glasgow R E, Vogt T M and Boles S M 1999 Evaluating the public health impact of health promotion interventions: the RE-AIM framework *Am. J. Public Health* **89** 1322–7
- Guruswamy L 2015 *International Energy and Poverty: The Emerging Contours* (Abingdon: Routledge)

- Hobbs R J and Norton D A 1996 Towards a conceptual framework for restoration ecology *Restor. Ecol.* **4** 93–110
- IEA, IRENA, UNSD, World Bank 2021 *Tracking SDG 7: The Energy Progress Report* (Washington, DC) (available at: [www.iea.org/reports/sdg7-data-and-projections/access-to-clean-cooking](https://www.iea.org/reports/sdg7-data-and-projections/access-to-clean-cooking)) (Accessed 17 January 2022)
- ISO TC 285 2018 Harmonized laboratory test protocols, part 1: standard test sequence for emissions and performance, safety and durability
- Jagadish A and Dwivedi P 2018 In the hearth, on the mind: cultural consensus on fuelwood and cookstoves in the middle Himalayas of India *Energy Res. Soc. Sci.* **37** 44–51
- Jeuland M A and Pattanayak S K 2012 Benefits and costs of improved cookstoves: assessing the implications of variability in health, forest and climate impacts *PLoS One* **7** e30338
- Kodros J K, Carter E, Brauer M, Volckens J, Bilsback K R, L'Orange C, Johnson M and Pierce J R 2018 Quantifying the contribution to uncertainty in mortality attributed to household, ambient, and joint exposure to PM<sub>2.5</sub> from residential solid fuel use *GeoHealth* **2** 25–39
- Koestler A 1970 Beyond atomism and holism—the concept of the holon *Perspect. Biol. Med.* **13** 131–54
- Lacey F G, Henze D K, Lee C J, van Donkelaar A and Martin R V 2017 Transient climate and ambient health impacts due to national solid fuel cookstove emissions *Proc. Natl Acad. Sci.* **114** 1269–74
- Makonese T, Pemberton-Pigott C, Robinson J, Kimemia D and Annegarn H 2012 Performance evaluation and emission characterisation of three kerosene stoves using a heterogeneous stove testing protocol (HTP) *Energy Sustain. Dev.* **16** 344–51
- Malhotra P, Neudoerffer R C and Dutta S 2004 A participatory process for designing cooking energy programmes with women *Biomass Bioenergy* **26** 147–69
- Medina P, Berrueta V, Martínez M, Ruiz V, Ruiz-Mercado I and Masera O R 2017 Closing the gap between lab and field cookstove tests: benefits of multi-pot and sequencing cooking tasks through controlled burning cycles *Energy Sustain. Dev.* **41** 106–11
- Morgan M G, Henrion M and Small M 1990 *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge: Cambridge University Press)
- Naeher L P, Brauer M, Lipsett M, Zelikoff J T, Simpson C D, Koenig J Q and Smith K R 2007 Woodsmoke health effects: a review *Inhal. Toxicol.* **19** 67–106
- Quinn A K, Neta G, Sturke R, Olopade C O, Pollard S L, Sherr K and Rosenthal J P 2019 Adapting and operationalizing the RE-AIM framework for implementation science in environmental health: clean fuel cooking programs in low resource countries *Front. Public Health* **7** 389
- Quinn A, Bruce N, Puzzolo E, Dickinson K, Sturke R, Jack D W, Mehta S, Shankar A, Sherr K and Rosenthal J 2018 An analysis of efforts to scale up clean household energy for cooking around the world *Energy Sustain. Dev.* **46** 1–10
- Sagar A D 2005 Alleviating energy poverty for the world's poor *Energy Policy* **33** 1367–72
- Seuring S and Müller M 2008 From a literature review to a conceptual framework for sustainable supply chain management *J. Clean. Prod.* **16** 1699–710
- Shiell A, Hawe P and Gold L 2008 Complex interventions or complex systems? Implications for health economic evaluation *BMJ* **336** 1281–3
- Steckler A and McLeroy K R 2008 The importance of external validity *Am. J. Public Health* **98** 9–10
- Stoner O, Lewis J, Martínez I L, Gumy S, Economou T and Adair-Rohani H 2021 Household cooking fuel estimates at global and country level for 1990–2030 *Nat. Commun.* **12** 5793
- United Nations 2021 Goal 7: ensure access to affordable, reliable, and sustainable energy for all (available at: <https://sdgs.un.org/goals/goal7>)
- Visser P 2005 The testing of cookstoves: data of water-boiling tests as a basis to calculate fuel consumption *Energy Sustain. Dev.* **9** 16–24
- World Health Organization 2018 Clean household energy solutions toolkit (available at: [www.who.int/tools/clean-household-energy-solutions-toolkit](http://www.who.int/tools/clean-household-energy-solutions-toolkit))