

# Monitoring and evaluation of improved biomass cookstove programs for indoor air quality and stove performance: conclusions from the Household Energy and Health Project

**Kirk R. Smith**

*School of Public Health, University of California, Berkeley, CA 94720-7360, USA  
E-mail: krksmith@berkeley.edu*

**Karabi Dutta**

*Appropriate Rural Technology Institute, 2nd Floor, Manini Apartments, Dhayarigaon, Pune 411041, India*

**Chaya Chengappa**

*Development Alternatives, 23, Sree Rama Mandir Road, Basavanagudi, Bangalore 560004, India*

**P.P.S. Gusain**

*Development Alternatives, 111/9-Z Kishangarh, Vasant Kunj, New Delhi 110070, India*

**Omar Masera and Victor Berrueta**

*Grupo Interdisciplinario de Tecnología Rural Apropiada and Centro de Investigaciones en Ecosistemas Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro No. 8701 Colonia Ex-Hacienda de San José de la Huerta, 58190, Morelia, Michoacán, Mexico*

**Rufus Edwards**

*Department of Epidemiology, School of Medicine, University of California, Irvine, CA 92697-3957, USA*

**Rob Bailis**

*Yale School of Forestry and Environmental Studies, 205 Prospect St., New Haven, CT 06511, USA*

**Kyra Naumoff Shields**

*School of Public Health, University of California, Berkeley, CA 94720-7360, USA*

*Standardized techniques for monitoring and evaluating (M&E) changes in indoor air quality and stove fuel performance were developed and deployed in two NGO-led programs to disseminate improved cookstoves (ICSs) in India and one in Mexico. This paper describes the objectives and characteristics of these monitoring and evaluation methods and how they were deployed.*

*The results showed major and mostly statistically significant improvements in 48-hour indoor air pollution concentrations in those households using the stoves one year after introduction. Kitchen levels of carbon monoxide reduced 30-70 % and concentrations of small particles reduced 25-65 %.*

*Results for stove performance were mixed, with some stoves achieving improvement in one or another of the short-term metrics that are part of the water boiling test (WBT) used to evaluate stoves in laboratory (controlled) settings. The kitchen performance test, which measures fuel use in households under actual use, was less easily conducted because of high variation and difficult field logistics. The results are more promising, however, with statistically significant reductions in fuel use per person ranging from about 20 to 67 %. From the results, it also seems clear that several indicators of stove performance derived from the WBT are not good predictors of actual fuel use and thus should be confined to evaluations during the design stage of stove development.*

*In two of the sites, the reductions in pollution roughly matched those in fuel use, although in the third, indoor air pollution may have reduced a bit more. This indicates perhaps that for all the monitored stoves, much or all of the benefits of each type came from improving the heat transfer into the pots and not from either increased combustion efficiency of the fires or stove-venting*

*(reliably working chimneys). More analyses are planned to explore these and other aspects of the stoves.*

*A range of recommendations are provided for future M&E efforts, with the primary one being to combine efficacy tests (small number of carefully monitored households under normal conditions) with larger well-designed surveys (questionnaire only) to determine actual usage and household perception. It is recommended that only those NGOs planning to develop significant long-term capability in measuring air pollution and stove performance under field conditions be expected to undertake effectiveness testing, i.e., evaluate population-wide changes from real large-scale dissemination programs. The alternative is to employ professional survey and environmental consulting firms, which also has the advantage of assuring independence of the process. In either case, over the long run it is important to generate national capacities for this kind of work.*

## 1. Introduction

Among the earliest and certainly the two largest and longest programs for introducing improved cookstoves (ICSs), the Indian National Programme for Improved Chulhas (NPIC) and the Chinese National Improved Stove Program (NISP) ran from 1983 to approximately 2002 and 1996 respectively. They are credited with introducing approximately 210 million stoves between them, 85 % in China, and thus affecting the lives of more than a billion people. Although there were great local variations in both countries, most observers believe that NPIC was able to achieve neither a significant sustained improvement in fuel efficiency, its major goal, nor its secondary objective of saved time and reduction of deforestation [Kishore and Ramana, 2002; TERI, 1989]. There was also little evidence that another secondary objective, improving indoor air quality, was widely achieved [Smith et al., 1983; Ramakrishna et al., 1989; Saksena et al., 2003]. NISP, on the other hand, is believed to have achieved major portions of its goals and has been termed one of the most successful Chinese programs in both energy efficiency and rural development. Sustained reductions in fuel use and air pollution were documented, although there is now need for dissemination of more advanced biomass fuel technologies to meet modern expectations [Sinton et al., 2004].

Why was there such a big difference in program success? Although there are many factors that help explain it, among the most important was that NISP included an explicit component of monitoring and evaluation or M&E (“checking”) in its operation, while NPIC did not. The national agency responsible for NISP – the Ministry of Agriculture – contracted directly with county-level rural energy offices to support their ICS activities. When a county claimed to have met its goals, province-level rural energy offices were engaged to verify that sufficient stoves were actually in place (usually 85 %), had been built to specifications (design and materials), and met minimum fuel efficiency requirements (using a simplified water boiling test, WBT). This was done through a simple randomized survey in which households were actually visited by the provincial team. Counties did not receive their final payments from the Ministry of Agriculture until this “checking” had been done [Smith et al., 1993].

Today, the perceived value of improved energy efficiency and reduced smoke exposures in solid-fuel-using households is even greater than in the past due to much

more complete information about health effects and increased concerns about greenhouse gas (GHG) emissions in addition to the still relevant issues related to household economy and time use and local biodiversity protection in some areas. Although there are unlikely to be repeats of such large “top-down” ICS programs of the types initiated in India and China in the 1980s, there is a growing interest to find ways to trigger more market-based approaches to foster the introduction of many tens of millions of improved stoves, perhaps combined with improved fuels in some areas.

Whether intending to achieve economic, health, environmental, or climate protection goals or a combination, however, there is growing understanding of the need to verify what programs actually achieve in order to justify the expenditures by donors and investors. There are too many seemingly worthy actions to be taken and too few resources available to ignore the need to compare the cost-effectiveness of the alternatives. How much carbon is kept out of the atmosphere? How much is child mortality reduced? How much women’s time is saved? How much less do families pay for fuel?

Recognizing the high value of having standard M&E methods for ICS programs, the Shell Foundation (SF) provided a grant to the Indoor Air Pollution Group of the School of Public Health at the University of California, Berkeley, to coordinate the development and testing of methods for evaluating changes in indoor air quality (IAQ) and stove performance (SP), the Household Energy and Health (HEH) Project. These methods were deployed by three NGOs that had ICS programs separately funded by SF grants. This and the five papers that follow in this special issue of *Energy for Sustainable Development* (ESD) present the results of these efforts.

### *1.1. History of previous monitoring and evaluation (M&E) efforts*

In the 1980s, several tests of the fuel-related performance of biomass cookstoves were developed, some of which have occasionally been deployed as part of programs to disseminate ICSs [Bailis et al., 2007, in this issue]. Little development or formal evaluation of these methods has occurred since, however. In recent years, there has also been growing interest in disseminating ICSs to reliably improve IAQ by use of chimneys, improved combustion, and other techniques. Although measuring IAQ in biomass-using households has been the subject of numerous research studies in the last few decades [Saksena et

al., 2003], standard monitoring methods that could be deployed by non-research groups, such as NGOs, interested in evaluating ICS programs have not been developed.

### 1.2. M&E strategy for the HEH Project

To understand and evaluate the M&E methods deployed in this project, it is important to understand the philosophy behind the choices made at the start in designing the M&E strategy.

M&E of the impacts of a household intervention, such as an ICS, can occur at several levels.

1. *Controlled tests.* In laboratory or near-laboratory settings with simulated cooking, these are easiest, quickest, and cheapest to conduct, but reveal the technical performance of a stove, not necessarily what it can achieve in real household settings, particularly with devices like stoves that are sensitive to operator behavior and use fuels that vary in composition, moisture, and size.
2. *Efficacy tests.* These are conducted in real households cooking real meals, but under controlled settings in which every effort possible is made to ensure the stove is used to its best effect. They reveal what is maximally possible in households, but not necessarily what is actually achieved.
3. *Effectiveness tests.* These are conducted in the course of an actual dissemination effort with real populations cooking normally and give the best indication of real-world changes. Only these can determine actual usage under realistic promotion schemes. Evaluating such trials is more expensive, difficult, and time-consuming, partly because high variability is in the nature of real households and thus sample sizes must be fairly large to obtain statistically significant results.

As the original impetus behind the Shell Foundation M&E Project was to understand the impact on populations, it was decided that the focus would be on effectiveness, i.e., M&E as much as possible aimed to determine the effects in real populations. For this reason, effectiveness measures were chosen for each outcome.

- *Stove performance.* The kitchen performance test (KPT), an effectiveness test, was chosen as the primary measure, which involves measurements of fuel use in real households over several days. In addition, given its common use and relative simplicity, improved protocols were also developed and training provided for the WBT, a controlled test. An updated controlled cooking test (CCT) was developed as well, but only used by one group. (See [Bailis et al., 2007] in this issue for details.)<sup>[1]</sup>
- *Indoor air quality.* Using a longitudinal “before and after” design, a field survey of IAQ levels in actual households that were part of the stove disseminations was developed and implemented. (See [Edwards et al., 2007] in this issue for details.)

A number of other decisions were also made at the outset that had important implications for conducting the M&E activities. Among the most important are the following.

- The Project focused on IAQ and SP M&E activities that would result in quantitative metrics meeting nor-

mal statistical conventions for making judgments, for example, having sufficient sample size to be able to say that there was no more than 5 % probability that results showing a measured IAQ improvement were obtained by chance. This required training in basic statistical concepts and analysis both for choosing sample sizes and conducting statistical tests with monitoring data. The reasoning for taking this admittedly challenging approach was that quantification of benefits that meet standard statistical criteria would be most effective in convincing donors, local authorities, and government agencies that significant benefits had accrued.

- Although personal pollution exposures are the best indicator of health impact, they are substantially more difficult to measure in household settings than kitchen concentrations, and introduce additional variability and ethical considerations. Since the objectives of the NGOs were to have some statistical evaluation of the performance of the stove in its most common usage, rather than introducing additional variability inherent in personal exposures due to variability in time spent in different locations, and to avoid asking the groups to undertake even larger and more difficult efforts in their first M&E program, it was decided to focus on stationary IAQ measurements in the kitchen. Although the percentage reductions in indoor levels may not be equivalent to reductions in personal exposures (due to time spent out of the kitchen), the metric provides a measure of the effectiveness of the ICS in reducing the high levels of smoke in the kitchen from open fires and other traditional cookstoves. This is also consistent with many dozens of other studies undertaken around the world, mostly for research purposes, that have used indoor measurements due to the complexity and high participant impact of collecting personal exposure samples.
- Both to reduce variability caused by season and to measure longer-term changes, not just those immediately after stove introduction, the main “before and after” measurements were undertaken a year apart. This length of time introduced difficulties, however, such as retention of households in the sample.
- Although there are many pollutants in biomass smoke that probably play a role in the wide range of health effects that have been measured or suspected, it was decided to focus on the two most measured pollutants – *small particles and carbon monoxide*. They are created by somewhat different processes during combustion and have different mechanisms of action in producing health effects. Thus, they probably cover a good part of the spectrum of toxicity of the smoke, but certainly not all, as biomass smoke contains hundreds of toxic materials [Naehler et al., 2007]. Reductions in both, however, can probably be viewed as a good indication that most other pollutants have been reduced as well.
- To reduce variation between measurements in the same house, a perennial problem with household studies, a 48-hour rather than the more commonly used 24-hour

measurement period was used, which complicates logistics somewhat.

Because capacity-building was a high priority, it was decided that all M&E activities would be undertaken by the NGO staff with equipment and methods that were sustainable, i.e., could be used by them after the end of the Project with normal maintenance.

Because there are no inexpensive reliable commercial devices available, the choice to measure small particles required one of the following alternatives:

- purchase of expensive commercial instruments, which could not be fit within the budget;
- temporary employment of outside instruments and experts to do the work, which did not meet the capacity-building goal; or
- use of the UCB Particle Monitor (UCB), a novel inexpensive instrument developed for research purposes under a previous SF grant, but showing high promise for use by non-research personnel, although not yet tried out in this role [Litton et al., 2004; Edwards, et al., 2006; Chowdhury et al., 2007]<sup>[2]</sup>.

The last approach was taken, which, as described below, led to more work than envisaged because of the additional development of the UCB system that was needed in software and data processing.

### 1.3. Introduction to the Household Energy and Health projects

This paper and the five that follow document efforts to help fill these gaps. Three NGOs – Appropriate Rural Technology Institute (ARTI), India; Development Alternatives (DA), India; and Grupo Interdisciplinario de Tecnología Rural Apropiada (GIRA), Mexico – received funding in 2004 to disseminate ICSs in rural areas. To facilitate evaluation of these efforts a team at the University of California at Berkeley (UCB) developed and deployed jointly with these groups a novel package of monitoring and evaluation (M&E) methods to systematically document the impact of each NGO's ICS dissemination on IAQ and SP as compared to traditional cookstoves (TCSs) used locally. Concurrently, a team at Liverpool University developed and deployed M&E methods for health and quality of life assessment in conjunction with these NGOs [Bruce et al., 2007].

This M&E package consisted of four physical elements:

1. written materials on principles of field sampling, the pollutants, measurement methods, sample size selection, home selection criteria, statistical analysis, and data handling;
2. written protocols for use of the instruments, sample field monitoring questionnaires, data collection forms and spreadsheet templates for entering and evaluating the data;
3. equipment kits for IAQ monitoring and SP testing, including computer software as well as instruments and associated supplies; and
4. a website for access by the NGOs to obtain updates of software, protocols, and other project materials, <http://ceihd.berkeley.edu/heh.IAPprotocols.htm>.

In addition to the physical elements, the NGOs were actively

supported through several training sessions in each country by UCB staff over the course of the project; the sessions were evaluated through anonymous questionnaire surveys that were collected and analyzed independently from the personnel performing the training<sup>[3]</sup>. The training visits generally occurred separately for IAQ monitoring and SP testing as, although personnel that conducted the M&E and the SP were separate, the intention was to focus activities around one methodology so supervisors could be familiar with all aspects of the project and not to burden the NGO too heavily at one time. Detailed descriptions of the training activities were provided in previous project reports to the SF. Training visits for the M&E included:

1. an initial training workshop;
2. on-site training of field teams and evaluation of initial deployment in homes;
3. data-cleaning and database compilation workshop;
4. data analysis workshop<sup>[4]</sup>; and
5. stove performance test training.

This paper focuses on an assessment of effectiveness of the M&E package in this application and summarizes the results for IAQ and SP found by the stove programs by use of the package. Details of the results are found in other papers in this issue of ESD.

## 2. Methods

Here we provide a generic description of the methods used by the programs to measure IAQ and SP. Modifications of the standard methods by the groups that were required because of local circumstances are described in the papers that follow. Actual protocols and data worksheets are found on the HEH website, now maintained by UC Berkeley's Center for Entrepreneurship in International Health and Development (CEIHD):

<http://ceihd.berkeley.edu/heh.IAPprotocols.htm>.

### 2.1. Study design

There are basically three field study designs for determining whether an ICS lowers IAP levels in households compared to the TCS.

1. *Cross-sectional*. Measure IAP and SP in both ICS and TCS houses in the same geographic area, but after the ICS has been introduced. This is easiest to do, but difficult to interpret since households that have chosen to buy ICSs are likely to be different from those that retain TCSs. In addition, it is never certain how well a cross-sectional study reflects actual changes over time, which is the main interest of the NGOs.
2. *Before and after*. Measure in the same households before and after the ICS has been introduced. Here, many of the household parameters remain the same over time, thus reducing variability and the sample size needed to reach statistical goals. Although this gives a better indication of the parameter of interest, induced change in the households, it risks being affected by seasonal and other changes that affect all households over time.
3. *Before and after with control group*. Measure in the same household before and after, but with a portion not receiving an ICS, i.e., still having the TCS. In this

design, any seasonal or other changes that are affecting the entire community can be determined and separated from the effect of the ICS introduction. Maintaining a true control group, however, raises logistical and ethical issues in these settings.

There are a number of other statistical and logistical considerations also involved in choosing one of the designs [Edwards et al., 2007, in this issue]. After the initial training, all three groups choose Design #2 as being the best trade-off between the need for accuracy and the logistical difficulties and additional sample size required by addition of a control group. To reduce the effect of seasonal changes and to allow for the ICS to “settle in”, all groups planned to conduct measurements one year apart, although some groups did additional measurements as well.

### 2.2. Household selection

A screening questionnaire was used by field staff upon their first visit to each household to ensure that the household was suitable for and amenable to the study. If the head of the household agreed to be involved in the study, the field staff administered a consent form at that time of the type commonly required by institutional committees tasked with protecting the safety and rights of human subjects in research. As with many rural areas where housing is not standardized, a wide range of different kitchen and stove configurations are encountered in the rural areas where the NGO stoves were deployed. The intent of the M&E was not to measure the effect of the ICS in all configurations, which would have required a much larger sample size and associated monitoring effort and cost, rather to measure the most common situation in these areas.

### 2.3. IAQ monitoring

Particulate matter (PM) and carbon monoxide (CO) were measured in the same households both before and after the introduction of the ICS. Sample sizes were chosen to reflect the type of statistical analysis appropriate with this design and observed variability from previous studies around the world plus suitable allowance for drop-outs, data loss, etc.

PM was measured using the UCB. This instrument is sensitive to particles of aerodynamic diameter less than approximately 2.5 microns, called fine PM or PM<sub>2.5</sub>, which is the size range thought to be most important for health. The monitors were produced and individually calibrated in the Indoor Air Pollution Laboratory at UC Berkeley before their use. The UCB recorded PM concentration every minute in its memory, which then is downloaded into a personal computer afterwards. CO was measured with a commercial CO logger (HOBO, Onset Computer Corp)<sup>[5]</sup>, which also recorded concentration every minute. The HOBOS were purchased new and calibrated at the Indoor Air Pollution Lab at UC Berkeley using standard CO gas. In the middle of the sampling campaign before the start of the “after” sampling, a collocation calibration check was conducted in the field.

Briefly, at each household, a UCB and HOBO were placed next to each other on the wall of the kitchen for 48 hours, using defined criteria. The two devices were co-located (placed next to each other) in a relatively safe

location to minimize the risk of interrupting normal household activities or being disturbed or damaged. At the end of each sampling session, a post-monitoring questionnaire was administered to the main cook of the household. The questionnaire documented cooking and other activities that may have affected the kitchen IAQ during the monitoring period. The protocols and questionnaire used in this study, along with the HOBO CO calibration check protocol and the sampling data forms, are found on CEIHD’s website:

<http://ceihd.berkeley.edu/heh.IAPprotocols.htm>.

### 2.4. Stove performance testing

The controlled/laboratory stove test (WBT) and field effectiveness measure (KPT) are based on earlier versions developed by a group of stove development experts convened by a US-based developmental NGO in 1982 [VITA, 1985] and further elaborated in a technical report on stoves published soon after [Baldwin, 1986]. Modifications to the tests derived for this project were arrived at after extensive testing conducted in partnership with the Aprovecho Research Center. In both cases, changes were made to both simplify and standardize the tests with the goal of ensuring comparability and replicability within and across different stove organizations. The details of the protocols are discussed in the article on SP tests [Bailis et al., 2007] included in this issue.

The primary intention of this summary paper is to evaluate the M&E package developed and deployed in this project. The actual IAQ and SP results achieved by the projects are described in the papers that follow. We do compare briefly the results across programs [Bailis et al., 2007, in this issue], but it should be emphasized that these comparisons should be considered as merely indicative and not definitive, for the following reasons.

- Cooking needs and practices, foods, household parameters, fuels, climate, and other factors affecting stove performance and air pollution were different across the study areas.
- There are a number of complexities that prevent one-to-one comparison, due, for example, to the use of multiple stoves and fuels in some households and changes in conditions in different measurement periods.
- In many respects these data are of mainly historical interest at this point because the programs and their stoves have evolved over the course of the project.

This last factor, of course, is partly a direct outcome of the M&E efforts made by the NGOs as they responded to the interim results seen from these M&E efforts and to field visits made by the M&E technical advisors. This illustrates the advantage of ongoing M&E for continuous improvement of ICS programs, one of its main purposes.

### 2.5. Modifications of protocols in practice

Each group found it necessary to adapt the standard protocols to fit its own circumstances, as described below.

#### 2.5.1. Appropriate Rural Technology Institute

*Village selection.* The village must not be very far from the ARTI office and must have reasonable access by road. The village head must give permission and most of the houses in these villages should have the typical housing

pattern common in that area.

*Household selection.* In the villages selected, a short household survey was conducted using a survey sheet which ARTI developed. The criteria for household selection included family size, presence of children, kitchen architecture, economic status, housing pattern, willingness to buy the ICS, and willingness to participate. This information was entered into an Access database from which a selection was made so that all the houses in a particular area (high- or low-rainfall) had the same structure and building material, the family size was more or less the same (nuclear families were selected), there were young children below the age of 5 years, the households were willing to buy the stove and otherwise participate, and they were currently using the same traditional stove and fuel.

The households selected were informed and their consents were taken. WBTs were conducted in some of these households in each village. Wherever they agreed to make their cookstoves available for 3 days, KPTs were also conducted in these homes. Fewer agreed to a 7-day test.

#### 2.5.2. Development Alternatives

The project was implemented in three clusters of villages in the Bundelkhand region near Jhansi in central India (Radhapur, Niwari and Thona). Households that had elected to buy an improved stove were selected and monitored for baseline data, just before the monsoon, and then during the monsoon, winter and summer seasons. (The monsoon season is roughly late June-September, winter is November to February, and summer is March to June with temperature peaking in May-June.) The selection of the household was based on an extensive survey conducted initially using a pre-monitoring questionnaire based on criteria of kitchen type, kitchen size and family size. The households were selected from the survey after prior approval.

#### 2.5.3. Grupo Interdisciplinario de Tecnología Rural

##### *Apropiada*

Villages and households were selected considering the sample already chosen for a separate health study being conducted locally in which young women and at least one child less than 3 years old were selected for a total of 600 households in 6 villages. For the IAQ monitoring, an initial sub-sample was selected from the 600 households. After baseline measurements, some of the initial families were not willing or not able to participate in the tests in follow-up visits, and the sample was completed with new families.

*KPT.* The wood used for the test came exclusively from the households' own stocks. For these reasons, in some cases the test did not last the whole seven days. However, not supplying the families with additional wood allowed measuring the fuel consumption in actual local circumstances. Measurement of LPG consumption was also included, and confirmed that households that use only fuelwood for cooking save more fuel than those using both LPG and fuelwood. Moreover, households with both fuels also reduced LPG consumption following the introduction of the improved wood stove. The reason is that

LPG is expensive and having access to a clean and efficient woodstove, households started to switch from LPG to fuelwood for some cooking tasks previously done on the LPG stove.

*WBT.* The Patsari cookstove, with a 52-cm wide pan adapted for tortilla-making, was not designed to handle the 5-liter vessel to boil water required by the standard protocol. The problem was that heat was lost by the surface of the *comal* (a hot metal or ceramic plate for cooking tortillas) not in contact with the vessel. For these reasons measurements were done at 90° C instead of the 94° C that is the temperature at which water boils at the altitude where these households are located in the Mexican highlands.

CCTs were also performed for the main cooking task within Mexican rural households, tortilla-making. Results are available in [Masera et al., 2005].

### 3. Summary IAQ monitoring results

Details of the stoves being disseminated by the groups are provided in the following papers [Masera et al., 2007; Dutta et al., 2007; Chengappa et al.; 2007] in this issue. In summary, the stove disseminated by GIRA in the highlands of Mexico uses trunk wood and major branches and the stoves in both groups in India, ARTI and DA, were disseminated in rural areas where cowdung cakes, wood, and crop residues are used. In addition to the fuel difference, there are, of course, considerable differences in the traditional foods cooked in Mexico and India, and the stove designs reflect these differences. For example, on average almost two hours per day were spent cooking tortillas in Mexico. This follows low-temperature cooking of the dough (*nixtamal*), which does not require much tending by the cook.

There were also considerable climatic differences between the rural areas in India. ARTI disseminated stoves in predominantly highland areas (although within its sample it had groups both in highland areas and in lower-lying surrounding areas). The DA stoves were disseminated in predominantly flat agricultural regions.

Although CO and PM were measured in the same location, it would be expected that the reductions as a result of installation of an ICS would not exactly agree, since they are produced by different combustion processes, and behave somewhat differently (gas versus particles). The reductions for CO and PM for all groups, however, were generally in good agreement, especially for larger sample sizes.

Part of the protocol was to choose housing types that represent the most common conditions in the area, but not work in those that are infrequently found, in an attempt to reduce the variability associated with household measurements. (See [Edwards et al., 2007] in this issue.) This strategy is important as the variability is critical in determining the number of homes that are required to demonstrate statistical significance and the sample size is a significant factor in the cost, difficulty, time, intrusion, personnel, and other requirements of the field work. In spite of these efforts, the variability in all groups remained

high as a result of different stove usage patterns. Since stove usage patterns are a dominant factor in variability, in future they should be addressed through more targeted screening questionnaires in addition to clearly distinguishing differences in housing design and family size.

In order to account for drop-out rates, migration, differences in stove usage and measurement fallibility, one-third more houses were recruited and monitored than would be required by sample size calculations. As a result of this intentionally conservative selection, the differences in IAQ results are statistically significant for all the projects. Even so, due to the diverse stove adoption patterns, including several traditional stoves in use in some areas, an even more conservative sample size is recommended for future projects. This would allow more detailed statistical analysis of the relative importance of the different household, stove, and fuel factors that affect IAQ in the households.

The IAQ summaries here refer to households in which the stoves were compared, i.e., traditional stoves in the “before” phase and ICSs in the “after” phase, and thus do not take into account those households that had stopped using the ICSs for one reason or another. They also do not indicate the presence of other combustion sources in the household environment, such as stoves for cooking animal fodder, which can influence results.

A brief summary of the primary IAQ results for the two pollutants in each of the three areas is presented in Tables 1-3. Many more details, analyses, and discussion are found in the three papers following in this issue. In general, the stoves seem to have achieved significant improvements of 48-hour kitchen concentrations ranging from one-third to two-thirds of the original levels in the same household using the local traditional cookstove a year earlier. Note that all reductions are statistically significant, i.e.  $p < 0.05$ . It is important to point out, however, that the absolute levels in the household before introduction of the ICS in each area were quite different, ranging from 0.52 mg/m<sup>3</sup> (milligrams of pollutant per cubic meter of air) to 1.25 mg/m<sup>3</sup> for PM<sub>2.5</sub>. This presumably reflects the initial variations in housing, cooking types, fuel management, and other characteristics of the different regions. Although not seen by comparison across regions, however, in general those households starting with higher levels tended to experience greater reductions. See details in the companion papers in this issue.

There is not enough room here to discuss the health implications of these reductions. In summary, on an absolute basis these results imply major reductions in air pollution concentrations breathed by rural people. They are equal to, or in the case of Mexico, larger than would be the case of changing the PM<sub>2.5</sub> level in the dirtiest large city in the world to that in the cleanest, i.e., perhaps 0.3 mg/m<sup>3</sup> to 0.007 mg/m<sup>3</sup> [Cohen et al., 2004]. Even considering that kitchen concentrations do not correspond exactly to levels breathed by people all day, this is obviously a major achievement. This is the good news.

The bad news, however, is that even the lower particle concentrations with improved cookstoves are still far

**Table 1. ARTI (all Laxmi stoves): improvements of kitchen concentrations**

	Before, with TCS (48-hr ave, mg/m <sup>3</sup> )	After, with ICS (48-hr ave, mg/m <sup>3</sup> )	Absolute reduction (mg/m <sup>3</sup> )	% reduction
PM <sub>2.5</sub>	1.25	0.94	0.31	25
CO	12.4	7.6	4.8	39

**Table 2. DA: improvements of kitchen concentrations**

	Before, with TCS (48-hr ave, mg/m <sup>3</sup> )	After, with ICS (48-hr ave, mg/m <sup>3</sup> )	Absolute reduction (mg/m <sup>3</sup> )	% reduction
PM <sub>2.5</sub>	0.52	0.33	0.29	37
CO	9.02	6.17	2.85	32

**Table 3. GIRA: improvements of kitchen concentrations**

	Before, with TCS (48-hr ave, mg/m <sup>3</sup> )	After, with ICS (48-hr ave, mg/m <sup>3</sup> )	Absolute reduction (mg/m <sup>3</sup> )	% reduction
PM <sub>2.5</sub>	1.02	0.34	0.68	67
CO	10.2	3.5	6.7	64

above World Health Organization (WHO) Air Quality Guidelines (AQGs) set to protect health [WHO, 2006]. The recently revised WHO Global AQG for PM<sub>2.5</sub> (Interim Target I) is set at 0.035 mg/m<sup>3</sup> for annual average concentrations, which is what is implied by these 48-hour average measurements in households for activities that are repeated daily. Even taking into account that biomass particles might be somewhat less hazardous than average urban pollution, a suspicion not borne out by the evidence [Naeher et al., 2007], and that kitchen levels are not breathed all day, there is still some distance to go before the particle levels in these biomass-burning households with ICSs are brought down to levels commensurate with global expectations for protecting health. Nevertheless, major reductions in health risk can be expected from the reductions achieved.

In contrast, however, although more difficult to interpret in terms of health, the kitchen CO levels achieved by the ICSs in the three regions did bring pollution levels to near the level recommended by WHO (10 mg/m<sup>3</sup> maximum for 8 hours [WHO, 2000]).

In Mexico, mean 48-h baseline open-fire PM<sub>2.5</sub> measurements (1.02 mg/m<sup>3</sup>) fall in the range of other Mexican studies [Saatkamp et al., 2000; Riojas-Rodriguez et al., 2001] but post-intervention measurements still remain 5 times higher than the National Ambient Standard for PM<sub>2.5</sub> of 0.065 mg/m<sup>3</sup> [Secretaría de Salud, 2003]<sup>[6]</sup>.

The concentrations found in this study fall within the range found in research studies worldwide [Saksena et al., 2003], as might be expected, but are notable for a number of reasons, including the following.

- Scientifically valid measurements were undertaken by NGO teams, in two of three groups unaffiliated with university research groups.

**Table 4. Percentage changes in performance of improved stove relative to traditional stove for selected indicators**

			GIRA		DA				ARTI							
			Patsari	Sukhad	Laxmi	Anandi	Laxmi	Bhagalaxmi	Grihalaxmi							
High-power test	Cold start	Time	253	***	61	*	81	**	63	**	147	***	29	NS	-16	**
		SC	195	***	-24	NS	1	NS	13	NS	-3	NS	-43	***	-51	***
		Eff	-55	***	51	**	14	NS	-6	NS	-38	**	5	NS	10	NS
	Hot start	Time	224	***	26	NS	35	*	-5	NS	60	***	6	NS	-25	**
		SC	39	NS	-12	NS	-36	NS	-14	NS	101	***	-7	NS	10	NS
		Eff	-6	NS	9	NS	74	**	7	NS	-48	***	4	NS	-9	NS
Low-power test	SC	-33	NS	-30	NS	-45	NS	58	*	24	**	7	NS	23	**	
	Eff	76	**	4	NS	4	NS	-10	NS	47	***	23	NS	27	**	

**Notes**

Cases in which ICS is significantly improved relative to the TCS ( $p \leq 0.10$ ) are shown as shaded entries. NS = not significant ( $p > 0.10$ ). The significances of other entries are: \*  $p = 0.10$ ; \*\*  $p = 0.05$ ; \*\*\*  $p = 0.01$ .  $p$  refers to the probability that the observed difference occurred by chance. By convention,  $p = 0.05$  indicates that a finding is statistically significant.

Time = time to boil; SC = specific fuel consumption; Eff = efficiency.

The Patsari is compared to the average of two different traditional stoves, which are described in [Masera et al., 2007] in this issue.

The Sukhad is described in [Chengappa et al., 2007] in this issue. DA's other stoves are variations in the same design: the Laxmi has two pot-holes, but lacks the split-level surface, and the Anandi only accommodates one pot.

ARTI's Laxmi, described in [Dutta et al., 2007] in this issue, is similar to DA's stove of the same name. The Bhagalaxmi and Grihalaxmi are variations on this design, but the Bhagalaxmi lacks a chimney.

- Application of novel and inexpensive monitoring devices achieved good results.
- Good study designs were applied, allowing statistically valid comparisons to be made.
- Measurements were taken over 48 hours, thus reducing the high natural variation that has tended to make interpretation of previous studies difficult.
- Measurements were taken a year after introduction of the stoves, not only reducing any effect of season, but demonstrating sustainable changes, not just those achieved shortly after dissemination of the ICS.

**4. Summary of stove performance test results**

As was discussed above, each NGO assessed stove performance using a WBT conducted in lab conditions and a KPT conducted under real household conditions. In addition, GIRA used a CCT, which incorporates elements of both the WBT and KPT. The results of SP testing are summarized briefly below. The SP protocols and results are described in detail later in this issue [Bailis et al., 2007].

**4.1. Water boiling test (WBT)**

The WBT includes a high-power (fast-boil) phase and a low-power (simmer) phase. Each phase yields a number of SP indicators including time to boil a fixed quantity of water, fuel consumption per unit of boiled water, and efficiency. Table 4 shows relative changes in SP between the ICS and each NGO's TCS for each of these indicators. Changes that are statistically significant ( $p = 0.10$ )<sup>[7]</sup> are shown by the shaded entries. For example, GIRA's Patsari stove took over 2.5 times longer to boil water than the TCS used for comparison, a difference that was highly significant ( $p = 0.01$ ), but the Patsari was significantly more efficient in the low-power phase ( $p = 0.05$ ).

Note that the majority of test results are not statistically significant. In addition, the WBT revealed no unambiguous

winners or losers among the stoves being promoted by the Shell HEH grantees. Those ICSs that performed better than the traditional competitor in some categories of tests also fared more poorly in others and even those that performed less well than the traditional competitor according to the majority of indicators still did better in other aspects.

Nearly every ICS took longer to boil 5 liters of water than its traditional counterparts. In some cases this was because the stove is designed for low-power cooking tasks, where high-power output is detrimental to the dish being cooked. In addition, as noted in the feedback to the tests, the volume of water used in these tests was inappropriate for some stoves. The increased boiling time is not a surprising result as all the stoves had high thermal masses, which take longer to heat up than lighter stoves or open fires. Such stoves tend to cook more slowly than an open fire when cold. All massive stoves, however, showed some improvement in hot-start conditions, as is evident from Table 4. Moreover, while the majority of statistically significant results in the high-power tests indicated that the ICSs did not perform as well as the TCSs to which they were compared, the majority of ICSs performed as well as or better than TCSs in the simmer tests.

**4.2. Kitchen performance tests (KPT)**

The KPT is used to demonstrate the effect of stove interventions on household fuel consumption. Daily measurements of fuel consumption are made in a representative sample of households using traditional stoves and ICSs over an extended period of time (3-7 days). Because testers actually enter people's kitchens and directly weigh the quantity of fuel consumed, the KPT is the preferred method of quantifying the ICS's impact on household fuel use. The KPT should also include a series of qualitative surveys conducted before and after dissemination of the

Table 5. DA's mean per capita daily energy consumption (MJ/day) based on three days of measurements

		Traditional stove			Sukhad			Comparison	
		n	Mean	SD	n	Mean	SD	% difference	p
All households		21	20.9	3.44	21	16.9	2.34	-19	< 0.001
Village	Gadkhini	7	20.6	2.62	7	16.4	2.42	-20	< 0.001
	Gundrai	7	21.5	2.82	7	17.3	2.36	-20	< 0.001
	Rajapur	7	20.7	4.62	7	17.0	2.25	-18	0.003
Season	Monsoon	21	20.7	2.31	21	16.5	1.47	-20	< 0.001
	Winter	21	23.8	1.54	21	19.2	1.18	-19	< 0.001
	Summer	21	18.3	3.59	21	15.0	1.86	-18	0.001

Notes

n = number of households; SD = standard deviation of the mean.

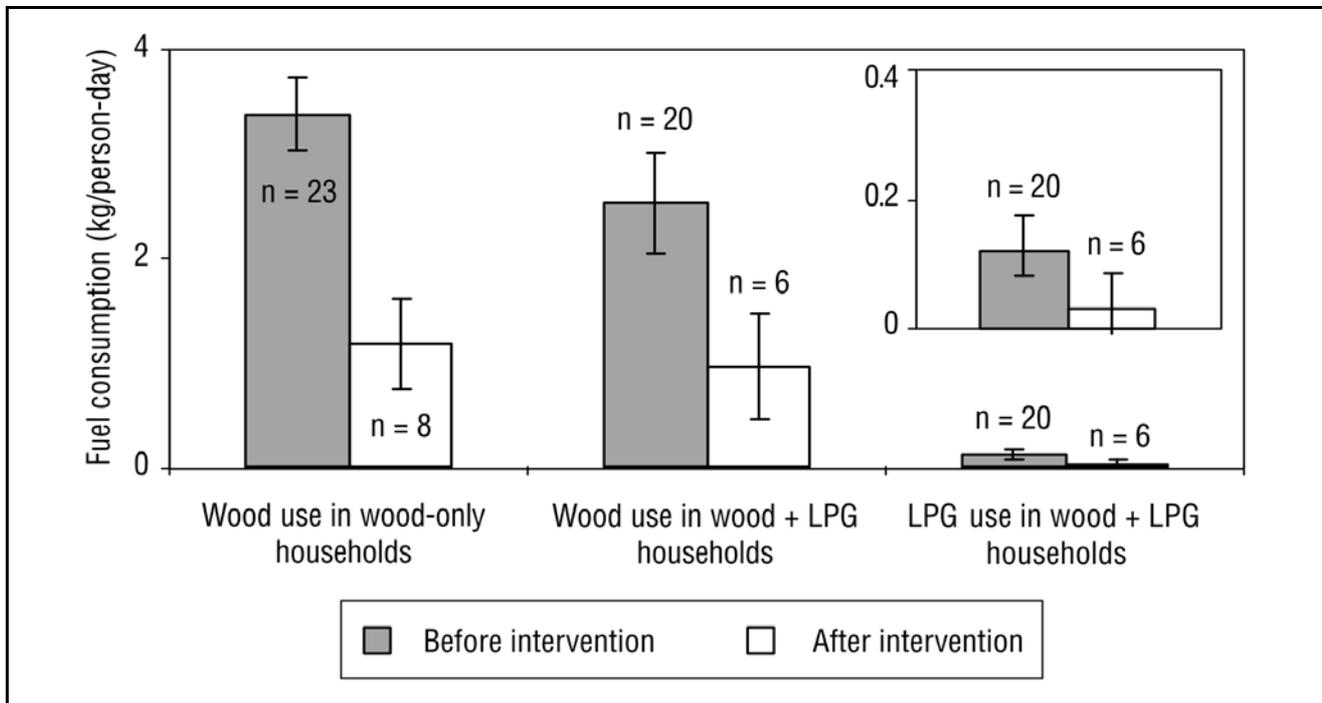


Figure 1. GIRA's daily fuel consumption in "wood-only" and "wood + LPG" households (the inset plot includes LPG consumption scaled up to show detail). Error bars show 95 % confidence intervals.

improved stove. This can capture non-quantitative aspects of household and community responses to the ICS. This paper discusses only the quantitative aspect of the KPT.

Although based on a simpler concept than the WBT, the KPT is more difficult for organizations to conduct in practice because of complicated sample selection procedures and logistical issues of working in real households. The former are particularly important because the variability in measurements of household fuel consumption tends to be higher than the variability observed in lab-based testing, hence larger sample sizes are needed to obtain statistically significant results. Several groups could not overcome this difficulty and followed different procedures.

One additional aspect of monitoring for stove performance that is associated with the KPT is the follow-up (post-intervention) qualitative surveys. Like field testing, this aspect of stove monitoring is logistically difficult and

may be overlooked by NGOs, particularly when resources are limited. This need is shared with that for IAQ studies. *Development Alternatives*. DA followed a before-after study design in which fuel consumption in seven participating households was measured daily for three days, yielding an average daily fuel consumption. All participants used the traditional stove and then the Sukhad stove. Some participants used both wood and dung in their stoves, but no fossil fuels were used for cooking during the KPTs. DA found reductions in fuel consumption consistently across villages and in all seasons. These findings are summarized in Table 5.

*Grupo Interdisciplinario de Tecnología Rural Apropiable*. GIRA utilized a paired-sample study design and followed the procedure for the KPT closely, measuring daily fuel consumption in each household for 5-7 days. It divided its sample into two groups to differentiate between households

**Table 6. ARTI's KPT results showing mean daily per capita energy consumption (MJ/day) among households using TCSs and households using either ICSs or FF stoves disaggregated by duration of the measurement period**

		Primary stove is TCS				Primary stove is ICS or FFS				Comparison	
		n		mean	SD	n		mean	SD	% difference	p
		days	HHs			days	HHs				
All measurement days		38	22	25.3	11.5	59	37	21.5	9.54	-15	0.09
Duration of KPT	One day	16	16	34.0	11.1	32	32	19.3	8.09	-43	< 0.001
	Three days	15	5	21.0	6.89	6	2	24.7	4.64	18	0.17
	One week	7	1	14.6	3.34	21	3	23.9	11.8	64	0.05

**Notes**

Households using ICS and FFS as their primary stove are grouped together because 12 out of 13 families listing FFS as their primary stove use an ICS as a secondary stove. n = number in sample; SD = standard deviation of the mean; HH = household.

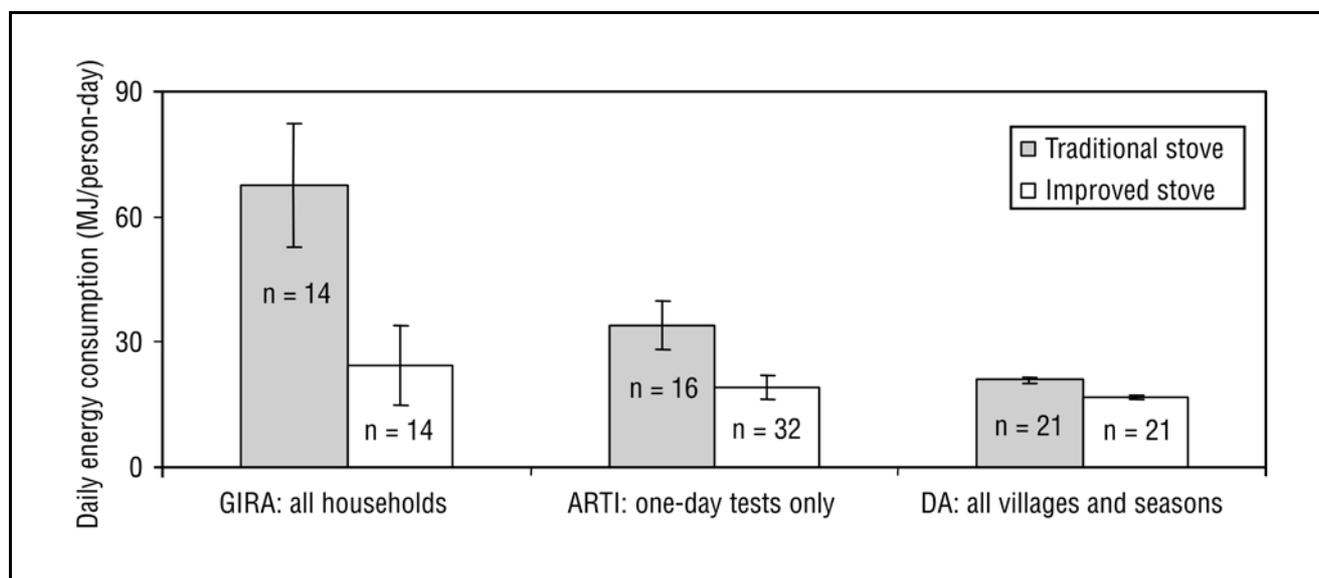


Figure 2. Average daily energy consumption in households measured by each NGO. All groups showed reductions in energy consumption that are statistically significant ( $p < 0.01$ ). ARTI's results only show households tested for one day.

that use both LPG and wood for cooking and households that only use wood [Berrueta et al., 2007]. Its results are summarized in Figure 1. The reductions in fuel consumption observed between the traditional stoves and Patsari stoves are highly significant ( $p = 0.01$ ). Interestingly, GIRA not only found that wood consumption and net energy consumption went down with the adoption of the Patsari stove, it also found LPG consumption decreased by 75 % among households using LPG ( $p = 0.01$ ).

*Appropriate Rural Technology Institute.* ARTI utilized a cross-sectional study design. Its findings require careful interpretation for several reasons. First, as with GIRA and DA, some participating households used multiple fuels and stoves, including wood and cowdung as well as kerosene and LPG. Only wood and dung consumption were measured, so the energy consumption ARTI finds fails to account for kerosene or LPG. However, ARTI reports that kerosene and LPG consumption was very low, even among families owning kerosene or LPG stoves. Hence, the error introduced by omitting fossil fuels from the analysis should be minimal. An additional complication arises because ARTI strayed from the KPT protocol by measuring households for different durations. In total,

ARTI collected 97 measurement-days of daily fuel consumption among 59 individual households. Of these, there were 48 households measured for one day each, 7 measured for three days, and 4 measured for one week.

Comparing fuel consumption among all measurement-days, we find a non-significant decrease of roughly 15 % in daily per capita energy consumption between households using a traditional stove and households using an ICS and a fossil-fuel (FF) stove. If the 48 households tested for one measurement-day are analyzed separately, the reduction in daily per capita energy consumption between households using a TCS and households using an ICS is 43 % and highly significant ( $p < 0.001$ ). In contrast, daily per capita energy consumption measured over a period of 3 or 7 days show that households with ICSs and FF stoves consume more wood per capita than households using traditional stoves. However, the differences are not significant due to the small sample size ( $n = 7$  and  $n = 4$  respectively). Daily per capita energy consumption and statistical comparisons are reported in Table 6.

Summing up, GIRA and DA observed statistically significant reductions in daily fuel consumption among households that adopted their respective ICSs. ARTI also

observed significant reductions in household fuel consumption, although the interpretation of its data is complicated by the variation in test days across its sample. The results of all NGOs' KPTs are illustrated in Figure 2.

#### 4.3. Comparison of lab-based and field-based test results

One additional goal of the SP testing was to assess the degree to which field performance is correlated to lab performance. Unfortunately, no clear pattern emerged in these studies. The lab-based WBT yielded a range of SP indicators: time to boil, specific consumption, and thermal efficiency for high-power tests; and specific consumption and thermal efficiency for low-power tests. Only specific consumption in the low-power simmer test has a correlation that exceeds 95 % confidence ( $r^2 = 0.83$ ,  $p \sim 0.01$ )<sup>[8]</sup>, indicating that higher (lower) fuel consumption during simmering in the lab is associated with higher (lower) fuel consumption in the field.

### 5. Summary and recommendations for future stove dissemination projects

We do not attempt an overall assessment of the achievements of the individual stoves here, which requires more detailed information that was not gathered as components of the IAQ and SP monitoring reported here. From these limited data, however, we do offer one observation. In two of the sites, the reductions in pollution roughly matched those in fuel use, although in the third, IAQ may have reduced a bit more. This might indicate that for some stoves much or all of the benefits of each type came from improving the heat transfer into the pots and not from increases in combustion efficiency of the fires nor in stove-venting (reliably working chimneys).

On the basis of the projects' experience in developing and implementing the M&E techniques described in this report, a series of recommendations for conducting M&E for IAQ and SP in future stove dissemination programs is presented below.

#### 5.1. IAQ monitoring recommendations

As noted in the introduction, unlike stove performance, M&E methods for establishing changes in IAQ have not been developed previously for ICS programs. Related methods used in research applications required significant modification, simplification, and standardization in this first effort and a number of lessons have been learned in the process.

- a. The project achieved its main objective in providing an estimate of the IAQ changes due to the ICSs for the three programs in a manner consistent with standard quantitative methods. In this the NGOs should be commended.
- b. The groups, however, exhibited important differences in the extent of their achievement. GIRA, which is associated with both national and international research groups, was able to accomplish most. It conducted significant monitoring beyond the scope of the HEH program itself, including assessment of personal exposures, measurements in other household locations, as well as measurements of volatile organic chemicals and greenhouse gases, and plans to continue this kind

of work. DA and ARTI conducted their HEH work well, but found it more challenging and resource-intensive than they expected, partly because of the initial difficulties with the UCB monitor.

- c. In retrospect, the decision to train NGO staff in simplified statistical theory and analysis methods may have been too ambitious. Although intelligent, motivated, and often well educated, NGO staff were more confident in such analyses when they had direct assistance from UC staff. An even more "cookbook" set of methods may have been more appropriate, although this approach risks misapplication when complex and unexpected field conditions are encountered, which is frequently the case.
- d. Household field studies to derive quantitative conclusions on elusive metrics such as air pollution are difficult to undertake, even by trained research groups. Thus, it is not surprising that NGO staff found it so as well. Part of the problem was a general underestimate of time and other resource requirements necessary for this kind of work, thus sometimes making the too few staff's work more difficult.
- e. Drop-out of households between sampling rounds plagued all the groups requiring care in choosing sample sizes with sufficient margins of safety and extra effort in motivating participation. Such losses occurred not only because households refused to continue, were constructing separate kitchens, had moved away, or otherwise became unavailable, but also because of significant continuation in the use of traditional stoves and transitional adoption patterns. Although this would be expected from social theories of technology dissemination within populations, this complicates collecting and analyzing the IAQ data in a consistent manner. Clearly, however, if improved adoption rates of the ICS are not seen over time, this would be a disappointing outcome of so much effort and resources to disseminate stoves.
- f. Multiple fuel and stove use in some areas complicated the collection and interpretation of changes in fuel use and air pollution. In such areas, future studies may have to do more careful stratification of households to obtain the statistical power to make judgments.
- g. Although in some ways the best for showing actual changes, one drawback of the before-after study design is that it is difficult to implement in a market-based dissemination effort, in which who will buy cannot be easily predicted and the motivation to participate is lower than with other types of dissemination. In such cases, a cross-sectional design may be more feasible, if less satisfying statistically and requiring greater sample numbers.
- h. Referring back to the choice of three principal measurement types, even though the best indicators of actual changes in the population, the "effectiveness" field-based measures deployed in the HEH Project may not be suitable for many NGOs in the future because of the difficulties noted above. There will be exceptions, of course, such as illustrated by GIRA.

- i. Using lab tests for IAQ assessments has not been validated and would seem to suffer from all those drawbacks related to SPTs (see below) plus others. In particular, there seems no way to realistically simulate either the leakage of smoke into a household from normal use of a chimney stove or the re-entry of outside smoke from the chimney back into the house. Research support for validating the relationships among IAQ measurements in the different settings (lab, efficacy, and effectiveness) is needed<sup>[9]</sup>.
- j. There is, however, an intermediate “efficacy” test that could be deployed that might be an appropriate alternative in many circumstances. This is the “test house” approach in which one or at most a few houses rented for the purpose are monitored as they change from no cooking, to traditional stove use, to ICS use, but using real cooks and cooking. It essentially reveals what the maximum improvement might be, but does not directly reveal what will be achieved on a population basis. It has also been recently used in a study in China after training by the UC Berkeley group [Xia and Chowdhury, 2006]. More description of this method is found in [Naeher et al., 2000].
- k. As planned, the two types of IAQ instruments (UCB for particles, HOBO for CO) were easily deployed in the field by the NGOs after training. They also held up well under field conditions. Much development was required, however, as these are the first projects globally where such an approach has been tried. With subsequent refinement of the techniques, subsequent projects implemented by CEIHD with somewhat simpler protocols have completed the assessment with a timeline similar to that initially estimated [Pennise with EWG, 2006].
- l. Because both instruments produce many data (minute averages over 48 hours), however, and require some manipulation of their outputs to produce reliable concentration measurements, data handling, cleaning, processing, and interpretation took much longer than anticipated, necessitating extra training sessions and more extensive involvement by the UC team than planned. These requirements were reduced late in the project to some extent by further development of the software to facilitate batch processing of files, but could not be entirely automated by the very nature of such continuous monitoring.
- m. Although there are no commercial alternatives that meet the cost criteria of the UCB, the UCB is a prototype monitor, which underwent significant development during the course of these HEH projects. While the intent was for the NGOs to input summary values into a database, in practice UC Berkeley’s direct participation was required in additional checking of data files, analysis, data processing, and data cleaning. Future efforts are unlikely to be hampered by these developmental concerns, however, but nevertheless do need to reserve resources for data handling and analysis.
- n. Future studies may wish to deploy devices designed to give summary values only, thus reducing the need

for data handling and processing. Although providing only a small fraction of the information available from continuous monitoring (nothing about peaks or temporal distributions, for example), simple means may be sufficient for many purposes of interest to NGOs. At present, unfortunately, although there are simple devices for giving means for CO concentrations, no easily deployed devices exist for particles.

Given the experience of this project, therefore, we have the following major recommendations for future large-scale NGO-driven ICS dissemination efforts.

1. Consideration should be given to relying principally on efficacy measures in the form of test-house studies in typical households. These studies can be done using outside professional monitoring teams from universities, research institutes, or environmental consulting firms, if the NGO is not interested in developing this capacity on its own. In this way, more expensive and reliable equipment can be used, because highly-trained people will be deploying it for short periods only, after which it can be used elsewhere.
2. Field surveys of pilot interventions are necessary, however, but can focus on assessing:
  - a. the degree to which the ICSs are actually put to use;
  - b. the extent to which traditional stoves remain in use (with or without the ICS);
  - c. the durability of the ICS in real kitchens; and
  - d. householder suggestions for improvements to the design or construction of the stove to maintain good IAQ, which may only be revealed after extended use (perhaps best revealed by focus group discussions).
3. Large-scale effectiveness monitoring efforts should only be undertaken with appropriate resources, training, and partners and after results from the above procedures indicate that serious population benefits may be occurring.

#### 5.2. *SP test recommendations*

Although lab tests failed to show significant improvement in stove performance between TCS and ICS, the ICSs promoted by these NGOs appear to reduce fuel consumption in actual use by between 19 and 66 %. In order to fully assess the impacts of these interventions, however, additional follow-up should be conducted in order to assess rates of stove adoption and stove degradation in the long term<sup>[10]</sup>.

In addition, stove adoption and use are dependent on many other factors including climate, cultural norms, and specific cooking needs. Since success of an ICS program is defined by the numbers of stoves in actual use in communities and the extent of multiple fuel and stove use, rather than simply the number of stoves that are disseminated, both adoption and performance within the household should be incorporated in evaluation efforts. For large-scale ICS projects, we have the following major recommendations, which are discussed in more detail in [Bailis et al., 2007] later in this issue.

1. The WBT should be utilized during the design stage of the ICS and perhaps to verify that stoves in place

Table 7. Comparison of testing methods

	Lab	Efficacy	Effectiveness
SP example	Water boiling test	Controlled cooking test	Kitchen performance test
IAQ example	None developed	Test house	Paired before-after field measurements <sup>[1]</sup>
Utility	Good for design phase stove development	Reveal best achievable in real households	Achievable in real household dissemination
Control	All factors except stove type	Household, fuel, food, cook, season, compliance	None, or reduce variability minimally by eliminating outliers in population
Advantages	Quick, comparable internationally, relatively simple if equipment available	Relatively quick, can be done by short-term team and thus use expensive equipment	Only way to provide indication of changes achieved by an intervention in real populations. Allows measurement of effect over time as well.
Disadvantages	Difficult to translate to other settings, weak connection to real performance in populations	Cannot be translated between populations, does not measure effect of compliance	Expensive, time-consuming, and requires relatively sophisticated field research skills

Note

1. Other field study designs are also possible, such as cross-sectional, prospective cohort with control group, etc. All share the same characteristics described here to a considerable degree.

2. The CCT, an *efficacy test* as defined above, should be promoted as a more appropriate and locally relevant measure of lab-based stove performance.
3. Field surveys are also necessary, in order to assess:
  - a. the degree to which the ICSs are actually adopted by the target population;
  - b. the extent to which traditional stoves remain in use (with or without the ICS);
  - c. the durability of the ICS in real kitchens; and
  - d. consumer suggestions for improvements to the design or construction of the stove.
4. Quantification of fuel consumption in the field (effectiveness surveys) is absolutely necessary if the agency implementing or funding the stove intervention wants specific data about the impact of the ICS on community, regional or national-scale energy consumption. Lab-based tests are not reliable predictors of fuel consumption in the field. Field assessments, however, should only be conducted if appropriate human and financial resources are available.

5.3. Final recommendations

For large-scale ICS-driven disseminations, therefore, we have the following major recommendations.

1. Consideration should be given to relying principally on efficacy measures in the form of controlled cooking tests in typical households. These studies can be done by NGOs after modest training, but short-term assistance in data analysis and interpretation may be needed from outside professional consultants from universities, research institutes, or environmental consulting firms, if the NGO is not interested in developing this capacity on its own.
2. As with IAQ monitoring, however, field surveys of pilot interventions will also be necessary, however, but can focus on assessing<sup>[11]</sup>:
  - a. the degree to which the ICSs are actually put to use in the population;

- b. the extent to which traditional stoves remain in use (with or without the ICS);
  - c. the durability of the ICS in real kitchens; and
  - d. householder suggestions for improvements to the design or construction of the stove to maintain good performance, which may only be revealed after extended use (perhaps best revealed by focus group discussions).
3. Large-scale effectiveness monitoring efforts using KPTs should only be undertaken with appropriate resources and training and after results from the above procedures indicate that serious population benefits may be occurring. For this reason, stove organizations need technical and financial support to do reliable and rigorous field monitoring. Donors must understand this and be willing to support these activities. Budgets for stove projects must reflect this need by including funds for in-group training, hiring an outside consultant, or subcontracting some or all M&E tasks to a third party organization.

In summary, Table 7 compares the three levels of tests – lab, efficacy, and effectiveness – for both IAQ and SP measurements, with brief mention of the advantages and disadvantages of each. Here, we recommend that future M&E efforts focus on the middle type, efficacy measures.

Finally, related to both IAQ monitoring and SP tests, credibility of results will always be enhanced by engaging independent organizations to conduct the M&E in order to reduce the chance of bias and the temptation to report partial results that do not fully reflect what was found. ■

Acknowledgements

We much appreciate input by Nigel Bruce, Dana Charron, Kristie Jagoe, David Pennise, and Dean Still, the leadership of A.K. Karve at ARTI and S. Patara at Development Alternatives, as well as the cooperation and patience of the householders in India and Mexico who participated in the surveys. Funding for this work was provided by the Shell Foundation, London. Acknowledgements for the work at the individual sites are found in the separate papers in this issue.

## Notes

1. For a recent review of stove testing methods covering developed and developing countries see [Kaisel, 2005].
2. It is difficult to predict the cost if the UCB were manufactured on a large scale, without any training, after-sales service, warranty, or profit and with no value assigned to the two software programs needed to operate it, but including pre-sales calibration and testing, however, a rough cost might be US\$ 250 each.
3. During the course of the project, Rufus Edwards, who was initially at the University of California, Berkeley, moved to the University of California, Irvine, where he continued his major role.
4. Some training also occurred in California.
5. [http://www.onsetcomp.com/solutions/products/loggers/\\_loggerviewer.php5?pid=228](http://www.onsetcomp.com/solutions/products/loggers/_loggerviewer.php5?pid=228).
6. India has no national standard for PM<sub>2.5</sub> yet.
7. *p* refers to the probability that the measured difference occurred just by chance, i.e., does not represent a true difference. By convention, if the *p*-value of a comparison is less than 0.05 (5 %) the difference is termed "significant".
8. *r*<sup>2</sup> is a statistical measure of the degree to which two measurements are correlated. In this case, it indicates that 83 % of the changes in the KPT results were predicted by the change in specific fuel consumption in the low-power simmer test.
9. As part of expanded research efforts triggered by the HEH Project, GIRA has conducted preliminary exploration of how well both laboratory WBTs and household WBTs perform in predicting emission. To date, analysis indicates little relationship. Emission data from real households, however, tend to agree with the KPT data in terms of wood savings reported.
10. For example, ARTI reported that many households in both the high- and low-rainfall regions stopped using their ICS after about 6 months. The problem was with the Laxmi stove in both areas. The region suffered a very heavy monsoon that year in both areas, causing the chimneys to leak and making the stoves damp, and, in some cases, the kitchen floor also became wet. Participants removed the Laxmi stoves and reinstalled their traditional stoves. In some cases they also demanded compensation for the roof and reimbursement for the cost of the stoves.
11. See also the companion study, [Bruce et al., 2007].

## References

- Ballis, R., Berrueta, V., Chengappa, C., Dutta, K., Edwards, R., Masera, O., Still, D., and Smith, K.R., 2007. "Performance testing for monitoring improved biomass stove interventions: experiences of the Household Energy and Health Project", *Energy for Sustainable Development, XI(2)* (this issue), pp. 57-70.
- Baldwin, S.F., 1986. *Biomass Stoves: Engineering Design, Development, and Dissemination*, Center for Energy and Environmental Studies, Princeton, NJ, 287 pp.
- Berrueta, V., Edwards, R.D., and Masera, O.R., 2007. "Energy performance of woodburning cookstoves in Michoacán, Mexico", *Renewable Energy* (forthcoming).
- Bruce, N.G., Jagoe, K., and Bromley, H., 2007. *Standard Monitoring Packages for Household Energy and Health Field Projects*, Reference no. 21084/21373, Department of Public Health, Liverpool University, January, <http://tinyurl.com/2x2hpu>.
- Chengappa, C., Edwards, R., Bajpai, R., Naumoff Shields, K., and Smith, K.R., 2007. "Impact of improved cookstoves on indoor air quality in the Bundelkhand region in India", *Energy for Sustainable Development, XI(2)* (this issue), pp. 33-44.
- Chowdhury, Z., Edwards, R., Johnson, M., Naumoff Shields, K., Allen, T., Canuz, E., and Smith, K.R., 2007. "An inexpensive light-scattering particle monitor: field validation", *Journal of Environmental Monitoring*, submitted.
- Cohen, A.J., Anderson, H.R., Ostro, B., Pandey, K.D., Krzyzanowski, M., Kuenzli, N., Gutschmidt, K., Pope, C.A., Romieu, I., Samet, J.M., and Smith, K.R., 2004. "Mortality impacts of urban air pollution", in Ezzati, M., Rodgers, A.D., Lopez, A.D., and Murray, C.J.L., (eds.), *Comparative Quantification of Health Risks: Global and Regional Burden of Disease due to Selected Major Risk Factors*, Vol. 2, World Health Organization, Geneva, pp.1353-1433.
- Dutta, K., Naumoff Shields, K., Edwards, R., and Smith, K.R., 2007. "Impact of improved biomass cookstoves on indoor air quality near Pune, India", *Energy for Sustainable Development, XI(2)* (this issue), pp. 19-32.
- Edwards, R., Smith, K.R., Kirby, B., Allen, T., Litton, C.D., and Hering, S., 2006. "An inexpensive dual-chamber particle monitor: laboratory characterization", *Journal of the Air and Waste Management Association*, 56, pp. 789-799.
- Edwards, R., Hubbard, A., Khalakina, A., Pennise, D., and Smith, K.R., 2007. "Design considerations for field studies of changes in indoor air pollution due to improved stoves", *Energy for Sustainable Development, XI(2)* (this issue), pp. 71-81.
- Kaisel, D., 2005. *International Review of Household Stove Standards for Performance and Emissions, Testing Methodologies, and Enterprise Innovation*, Center for Entrepreneurship in International Health and Development, University of California, Berkeley, <http://ceihd.berkeley.edu/Docs/Standards%20review%20final%20Dec%202020.pdf>
- Kishore, V.V.N., and Ramana, P.V., 2002. "Improved cookstoves in rural India: how improved are they? A critique of the perceived benefits from the National Programme on Improved Chulhas (NPIC)", *Energy, 27(1)*, pp. 47-63.
- Litton, C.D., Smith, K.R., Edwards, R., and Allen, T., 2004. "Combined optical and ionization measurement techniques for inexpensive characterization of micrometer and submicrometer aerosols", *Aerosol Science and Technology*, 38, pp. 1054-1062.
- Masera, O.R., Díaz, R., and Berrueta, V., 2005. "From cookstoves to cooking systems: the integrated program on sustainable household energy use in Mexico", *Energy for Sustainable Development, IX(1)*, pp. 25-36.
- Masera, O., Edwards, R., Armendáriz Arnez, C., Berrueta, V., Johnson, M., Rojas Bracho, L., Riojas-Rodríguez, H., and Smith, K.R., 2007. "Impact of Patsari improved cookstoves on indoor air quality in Michoacán, Mexico", *Energy for Sustainable Development, XI(2)* (this issue), pp. 45-56.
- Naehler, L.P., Leaderer, B.P., and Smith, K.R., 2000. "Particulate matter and carbon monoxide in highland Guatemala: indoor and outdoor levels from traditional and improved wood stoves and gas stoves", *Indoor Air*, 10, pp. 200-205.
- Naehler, L.P., Brauer, M., Lipsett, M., Zelikoff, J.T., Simpson, C., Koenig, J.Q., and Smith, K.R., 2007. "Woodsmoke health effects: a review", *J. of Inhalation Toxicology*, 19(1), pp. 1-47.
- Pennise, D., with EWG (EnterpriseWorks-Ghana), 2006. *Indoor Air Pollution Monitoring in Ghana: Third Progress Report*, Project no. 21341, submitted to the Shell Foundation, London, UK, April.
- Ramakrishna, J., Durgaprasad, M.B., and Smith, K.R., 1989. "Cooking in India: the impact of improved stoves on indoor air quality", *Environment International*, 15(1-6), pp. 341-352.
- Riojas-Rodríguez, H., Romano-Riquer, P., Santos-Burgoa, C., and Smith, K.R., 2001. "Household firewood use and the health of children and women of Indian communities in Chiapas, Mexico", *Int. J. Occup. Environ. Health*, 7(1), pp. 44-53.
- Saksena, S., Thompson, L., and Smith, K.R., 2003. *Database of Household Air Pollution Studies in Developing Countries*, Protection of the Human Environment, World Health Organization, Geneva, <http://ehs.sph.berkeley.edu/hem/page.asp?id=33>
- Saatkamp, B.D., Masera, O.R., and Kammen, D.M., 2000. "Energy and health transitions in development: fuel use, stove technology, and morbidity in Jarácuaro, México", *Energy for Sustainable Development, IV(2)*, pp. 7-16.
- Sinton, J.E., Smith, K.R., Peabody, J.W., Liu Y., Zhang X., Edwards, R., and Gan Q., 2004. "An assessment of programs to promote improved household stoves in China", *Energy for Sustainable Development, VIII(3)*, pp. 33-52.
- Smith, K.R., Aggarwal, A.L., and Dave, R.M., 1983. "Air pollution and rural biomass fuels in developing countries: a pilot village study in India and implications for research and policy", *Atmospheric Environment*, 17(11), pp. 2343-2362.
- Smith, K.R., Gu S.H., Huang K., and Qiu D.X., 1993. "One hundred million improved cookstoves in China: how was it done?", *World Development*, 21(6), pp. 941-961.
- Secretaría de Salud, 2003. *Proyecto de Modificación a la NOM-025-SSA1-1993*, Mexico City, Mexico.
- TERI (Tata Energy Research Institute), 1989. *Evaluation of Improved Stoves in Tamil Nadu, Rajasthan and West Bengal*, Tata Energy Research Institute reporting to the Department of Non-conventional Energy Sources, New Delhi.
- VITA (Volunteers in Technical Assistance), 1985. *Testing the Efficiency of Wood-Burning Cookstoves: Provisional International Standards*, Volunteers in Technical Assistance, Arlington, VA, USA.
- WHO (World Health Organization), 2000. *Air Quality Guidelines – 2000*, WHO Regional Office for Europe, Copenhagen.
- WHO (World Health Organization), 2006. *Air Quality Guidelines Global Update - 2005. Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide*, WHO Regional Office for Europe, Copenhagen, 484 pp., [http://www.euro.who.int/InformationSources/Publications/Catalogue/20070323\\_1](http://www.euro.who.int/InformationSources/Publications/Catalogue/20070323_1).
- Xia Z., and Chowdhury, Z., 2006. "IAP monitoring in Yunnan, China", presented at Partnership for Clean Indoor Air Meeting, Kirkland, WA, January 30, <http://www.pciaonline.org/assets/8-PCIA-IAP-Pres-UCB-CEIHD-TNC-China.pdf>