Gathering Time-Series Data for Evaluating Behavior-Change Campaigns in Developing Countries: Reactivity of Diaries and Interviews

Evaluation Review

34(5) 367-390

© The Author(s) 2010

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0193841X10383940

<http://erx.sagepub.com>

Robert Tobias¹ and Jennifer Inauen¹

Abstract

Gathering time-series data of behaviors and psychological variables is important to understand, guide, and evaluate behavior-change campaigns and other change processes. However, repeated measurement can affect the phenomena investigated, particularly frequent face-to-face interviews, which are often the only option in developing countries. This article presents three intervention control studies to investigate this issue. Daily diaries in Cuba did not affect behavior or attitudes for persons with

¹ Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Corresponding Author:

Robert Tobias, Eawag/Siam, Überlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland

Email: robert.tobias@eawag.ch

intervention but reduced attitudes for persons without intervention. Reactivity of face-to-face interviews in Bolivia was negligible if applied weekly, but strong if applied twice per week. The article concludes with recommendations for gathering time-series data in developing countries.

Keywords

reactivity of measurement, monitoring survey, time-series data, panel data, behavior-change campaign

Behavior-change campaigns are usually evaluated by comparing data gathered after interventions with data gathered from a group without intervention (e.g., Bamberg 2002; Cialdini, Reno, and Kallgren 1990) or with data gathered before the campaign (e.g., Bamberg and Schmidt 2001; Schultz et al. 2007). Such panel surveys provide valuable data for evaluating campaigns; however, relying only on data that are gathered every few months has a number of shortcomings: (a) Retrospective reports of behaviors, beliefs, feelings, and events might be biased by the memory processes involved to produce them; (b) In most cases, the processes initiated by campaigns develop over weeks, months, or even years, and the short-term effects can be very different from the long-term effects. Thus, the interpretation of the data from an arbitrarily set panel in relation to the ongoing process remains unclear; (c) A before–after perspective does not allow the investigation and thus understanding of dynamic (i.e., time-dependent) aspects of change processes. For example, two campaigns with similar effects might differ in that one reached this effect much faster; a campaign might have a smaller but more sustainable effect than an alternative campaign or a set of powerful interventions might fail just due to inadequate timing. To remedy these shortcomings, panel surveys should be combined with gathering data in short intervals, that is, the gathering of time-series data. Such data can then also be used to guide campaigns (e.g., to determine optimal moments for interventions or a panel wave) and to help better understand the determinants and “mechanics” of change processes.

The gathering of time-series data is a common practice in the social sciences (for an overview see Hoppe et al. 2000). Ideally, behavior is registered continuously in an unobtrusive way (e.g., telemetric television audience panels; Schmitz et al. 1996). However, to understand the situational effects and psychological processes determining the behavior, it is inevitable to ask the investigated persons questions. A number of approaches have been successfully implemented. Self-administered diaries are the least intrusive

method for gathering data in short intervals and work best for obtaining data on less frequently occurring behaviors, events, or beliefs and feelings that do not change too quickly (Hoppe et al. 2000). In the experience sampling method (ESM) participants are prompted by a pager at random times to provide data (Kubey, Larson, and Csikszentmihalyi 1996). This method is more intrusive than the diary method but leads to better results for frequent behaviors or events and rapidly changing thoughts or feelings. In addition, telephone interviews are used for frequent data gathering. Such interviews are quite obtrusive but have the advantage of reducing missing data and eventually improving data quality (Morrison, Leigh, and Gillmore 1999).

Gathering time-series data for evaluating campaigns in developing countries comprises additional challenges. In many regions, electronic equipment such as telephones or pagers are not available or cannot be used due to the lack of reception or electricity, danger of theft, or a lack of education to use such devices. Often, the high rate of illiteracy impedes any form of self-administered data gathering. In the worst case, data must be gathered by face-to-face interviews. This is a very obtrusive form of data gathering, which cannot be applied too frequently. Thus, the form of data gathering investigated in this article is not directly comparable to the methods mentioned above, where data are gathered daily or even more often with rather unobtrusive methods. Yet, there are important differences to traditional panel surveys. We will call this form of gathering time-series data *monitoring surveys*.

Monitoring surveys are panels of a few simple items that are applied frequently enough to capture the dynamics of change processes. The items have a short-term retrospective reference (e.g., "How often have you done x today?"). Thus, the reference is much less far in the past than in traditional panel surveys but neither addresses the current moment as in ESM (e.g., "Are you doing x right now?"). The questionnaires are shorter and simpler than those in most panel surveys but more extensive than those in ESM. There are many more measurement points than in traditional panel surveys (at least more than five, mostly several dozens), but the intervals of data gathering are larger than in ESM (daily to fortnightly). The methods of data gathering are the same used in panel surveys but additional measures might be applied to reduce memory biases. Thus, a monitoring survey might imply face-to-face interviews every few days. Consequently, reactivity effects become an issue here.

To investigate the effects of monitoring surveys on the data, we compared changes in behavior and attitudes of persons to which a monitoring survey was applied to those of persons without monitoring. This article

presents three studies that quantify the reactivity of gathering time-series data during behavior-change campaigns in developing countries. In the first study, data were gathered using self-administered diaries, whereas in the other two studies face-to-face interviews were applied at different intervals. In all three studies, the main barrier for performing the behavior was forgetting; thus, the interventions were based on memory aids and self-commitments. Since a reminder to perform the behavior is expected to be the strongest reactivity effect of monitoring surveys, these studies should give a rather conservative estimate of the reactivity of monitoring surveys. Since the focus of this article is on the reactivity of monitoring surveys, the time series gathered and their analysis is not presented here but published elsewhere (see Tobias 2009). Each study is presented in a section with its own methods and results part. In the concluding discussion of the article, the experiences from the three studies are compiled to give practical advice on how to perform a monitoring survey during a campaign in developing countries.

Study I: Promoting Recycling Behavior in Santiago de Cuba

Campaign and Method

The data of the first study were gathered at the beginning of 2005, in a campaign designed to reduce deposited solid waste in Santiago de Cuba (Binder and Mosler 2007; Mosler et al. 2006). On the basis of expert interviews and a representative survey that was conducted about a year before the beginning of the campaign (Mosler et al. 2008; Tobias, Bruegger, and Mosler 2009), the promotion of waste separation for recycling was identified as a possible way to reduce deposited waste. Although the attitudes toward waste separation had been positive, the performance was low because the participants forgot to separate their waste at the moment of disposal. A simple intervention based on reminders was consequently implemented. Every household in the sample received a sheet of paper with the Spanish text: "Please classify and separate! Glass—Aluminum—Paper—Cardboard—Plastic." Without further persuasion, the paper was handed out with the request for the participants to hang it up in the house at the location where most of the waste was collected.

The data were gathered during the pilot phase of the campaign, when the reminders were tested before being applied on a large scale. The sampling procedure is presented in Figure 1. Based on the aforementioned

Figure 1. Flow chart of the sampling procedure in Study I. CDR = Comité para la Defensa de la Revolución, the smallest political unit of the organization of Santiago de Cuba.

representative survey (Mosler et al. 2008; Tobias et al. 2009), two districts (repartos) were selected that were best suited for testing the intervention¹ and one as a control area. Within each district, a study location was randomly selected. The five neighboring CDRs (Comité para la Defensa de la Revolución, the smallest political unit of the organization of Santiago de Cuba) around this location formed the population from which the sample was drawn. The random route method was used for sampling (e.g., Hoffmeyer-Zlotnik 2003). The sample consisted of 176 households that received a reminder (intervention group) and 88 households without intervention (control group). Not all interviewers reported the refusal rate. Extrapolating from the refusal rate documented by some interviewers, it

was estimated to be about 10%. The interviews were performed with the person that is mostly responsible for waste disposal in each household. From the intervention group, data were gathered with a three-point panel: on the first day of the intervention (the interviewers distributed the reminders), after about a month later, and 1 month thereafter. From the control group, data were gathered only at the start of the intervention in the intervention group and 2 months later. For the panel survey, data were gathered through face-to-face interviews by local students, who had received an extensive training. Besides the panel survey, 64 households in the intervention group and 32 households in the control group were randomly assigned to be monitored daily during the first month of the campaign. Data were gathered using short questionnaires that the subjects filled out by themselves and which the interviewers collected approximately once a week. The questionnaires comprised 8 closed and 2 open items asking about the behavior (percentage of solid waste separated for recycling), attitudes, norms, and communication.

The questionnaires in the panel surveys comprised many more items. However, in this article, only three constructs will be discussed: self-reported behavior, affective attitude, and instrumental attitude. *Behavior* was measured by self-report, because in the survey prior to the intervention, measurement of observed behavior (i.e., analyzing the composition of solid waste not separated for recycling) was more problematic regarding reliability, validity, and practicability of the measurement. Constant biases due to social desirability are controlled for by the panel design, since changes over time subtract these. Behavior was measured with 10 items asking the percentage (0%, 25%, 50%, 75%, and 100%) of separated waste in the categories glass, aluminum, paper, cardboard, and plastic for the two institutions that recycle waste in Santiago de Cuba. The percentages separated within a category were added and the mean of these five sums forms the behavior measure of this study. Thus, the behavior scale ranged from 0 to 1. The *affective attitude* was measured using 2 items. One asked how much the persons feel like separating solid waste for recycling, and the other asked how pleasing or displeasing separating solid waste for recycling is. The first item had four categories (from 0 to 1) and the second had five (from -1 to 1). The values were added and divided by two. Thus, this scale had values from -1 to 1. In addition, the *instrumental attitude* was measured with 2 items. One asked, "How much do you think is it worth it *for you* to separate solid waste for recycling?" and the other asked, "How much do you think is it worth it *for the community and the environment* to separate solid waste for recycling?" Both items had seven categories

Table 1. Means, Standard Deviations, and Number of Cases in the Panel for the Groups of Study I

		Behavior		Affective Attitude		Instrumental Attitude	
		Monitoring	Panel	Monitoring	Panel	Monitoring	Panel
Control Panel 1	<i>M</i>	0.713	0.702	0.759	0.585	0.709	0.437
	<i>SD</i>	0.363	0.325	0.266	0.303	0.499	0.506
	Cases	20	46	20	45	20	45
Control Panel 3	<i>M</i>	0.583	0.596	0.362	0.552	0.458	0.596
	<i>SD</i>	0.371	0.288	0.326	0.355	0.422	0.399
	Cases	18	23	17	21	12	19
Intervention Panel 1	<i>M</i>	0.689	0.657	0.654	0.571	0.670	0.628
	<i>SD</i>	0.347	0.362	0.316	0.347	0.380	0.426
	Cases	56	112	56	111	57	111
Intervention Panel 2	<i>M</i>	0.838	0.770	0.687	0.631	0.718	0.670
	<i>SD</i>	0.218	0.296	0.318	0.300	0.356	0.412
	Cases	54	106	54	106	53	104
Intervention Panel 3	<i>M</i>	0.887	0.814	0.726	0.644	0.862	0.812
	<i>SD</i>	0.182	0.260	0.241	0.264	0.252	0.262
	Cases	54	107	54	107	54	107

ranging from -1 to 1 and the mean was used in the analyses. The scale ranged from -1 to 1 . The effects of the monitoring and the intervention on these three constructs were tested with the General Linear Model. Since all scales had the same range (the unipolar behavior scale from 0 to 1 and the bipolar attitude scales from -1 to 1), the effect sizes can be interpreted equally. Differences in the mean values smaller than 0.2 are interpreted as weak effects, between 0.2 and 0.4 as medium, between 0.4 and 0.6 as strong, and larger than 0.6 as very strong.

Results

The numbers of cases that entered the analyses are compiled in Table 1. The drop-out rate in the control group was considerably higher than in the intervention group. This dropout is, however, not caused by the refusal of the participants but by logistic problems in the second half of the campaign that led to many interviewers not visiting the participants. In Table 1, descriptive statistics are presented for the behavior and the two attitude components, divided into the three panel waves and the four investigated groups: with intervention versus without and with monitoring versus without. The means are further plotted in Figure 2.

Figure 2. Means of three variables at the three measurement times for the groups of Study I. Note that for cases without intervention no measurement is available for the second panel.

In Figure 2, a medium effect of the intervention on behavior and instrumental attitude can be observed, while monitoring does not seem to have an effect in the intervention group. However, a medium negative effect of the monitoring surveys on the attitudes can be observed for the control group. This could indicate participation fatigue and resentments of persons who did not understand why the data were gathered. This tendency is confirmed by analyses of variance presented in Table 2.

Overall, the results presented in Table 2 show a medium effect of the intervention on the behavior, a weak effect on the instrumental attitude, and no effect on the affective attitude. Monitoring did not have any effect at all. However, the following exceptions contradict this tendency: (a) in the intervention group, affective attitude is significantly higher for the monitored cases without affecting changes over time; (b) In the control group, the medium effect of monitoring on the affective attitude and on the instrumental attitude is apparent; (c) The effect of the intervention on behavior is stronger in the panel than in the monitoring group, while on the psychological constructs, it is stronger in the monitoring than in the panel group. The former might be explained by monitoring surveys' slight support of the effects of prompts, the latter is explained by the participation fatigue.

In this study, the reactivity of the monitoring—at least for the intervention group—is negligible and forms no problem. In the control group, negative effects of the monitoring on attitudes were observed (not, however, on

Table 2. Results of the Tests of Effects With the General Linear Model (GLM) in Study 1

Panel	Group	Effect	Test	Behavior		Affective Attitude		Instrumental Attitude	
				<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
1-3	<i>I</i>	<i>G</i>	<i>M</i>	1.513	.221	4.691*	.032	0.904	.343
1-3	<i>I</i>	<i>T</i>	<i>I</i>	30.157*	.000	3.664	.057	28.721*	.000
1-3	<i>I</i>	$T \times G$	$M \times I$	0.497	.482	0.001	.980	0.014	.905
1-3	<i>C</i>	<i>G</i>	<i>M</i>	0.383	.539	0.015	.903	1.242	.274
1-3	<i>C</i>	<i>T</i>	(<i>I</i>)	3.522	.068	13.385*	.001	0.962	.335
1-3	<i>C</i>	$T \times G$	$M \times (I)$	0.239	.628	12.379*	.001	5.750*	.023
1-3	<i>M</i>	$T \times G$	<i>I</i>	7.433*	.008	22.329*	.000	16.123*	.000
1-3	<i>P</i>	$T \times G$	<i>I</i>	12.819*	.000	0.534	.466	0.096	.757

Note. The first column indicates the measurement points considered, the second represents the groups. *I* represents cases with intervention and *C* (control group) represents cases without intervention. In both groups, cases with monitoring are contrasted with cases without monitoring. *M* stands for cases with monitoring and *P* (only panel) for cases without monitoring. In these two groups, cases with intervention are contrasted with cases without intervention. *Effect* indicates what effects are tested with the GLM: effects between groups (*G*), over time (*T*), or interaction effects of group with time ($T \times G$). The fourth column states what is tested in terms of this article: the effect of the monitoring (*M*), the intervention (*I*), or the interaction of monitoring and intervention ($M \times I$). Since the control group did not receive an intervention, the *I* is in parentheses. Results indicating significant effects are marked with an asterisk (*).

behavior). The cause for this effect might be fatigue of the participants that had to fill out the questionnaires without apparent reason, since they did not know about the interventions in other parts of the city. Thus, the traditional control group design may not be suitable when using monitoring surveys.

Study 2: Promoting Solar Water Disinfection (SODIS) in Cochabamba, Bolivia

Campaign and Method

In the second study, data were gathered during a campaign that was conducted in the city of Cochabamba, Bolivia, in fall 2005. The campaign promoted the use of SODIS to supply people with clean drinking water. SODIS is applied by filling water into transparent plastic bottles and laying them out in the sun for 6 hr (or 2 consecutive days if cloudiness exceeds 50%). This procedure inactivates pathogens that are responsible for water-borne diseases, such as diarrhea, dysentery, or cholera (Berney et al. 2006;

Joyce et al. 1996; Wegelin et al. 1994). For the households, SODIS is an economic way of disinfecting drinking water, while on a global level it reduces the burning of wood and gas used for disinfecting water by conventional boiling. In spite of its advantages, the dissemination of SODIS has not been as successful as expected. As research on the determinants of the use of SODIS in Bolivia indicates, weak habits and low attitudes seem to be major influences on the use of SODIS (Heri and Mosler 2008; Moser and Mosler 2008). However, attitudes were positive in the target population of the presented campaign. Therefore, prompts were implemented to prevent forgetting to execute the behavior and thereby facilitate the development of stronger habits for the use of SODIS. The prompts were designed to serve literate as well as illiterate persons. Posters (40 cm $\times$ 70 cm) combined the text “have you laid your bottles in the sun today?” with a picture of a girl drinking a glass of water and the five steps of how to apply SODIS (in words and pictograms). Interviewers distributed the prompts, and the time of the beginning of the intervention varied considerably between subjects. Furthermore, it should be noted that the control group was not geographically isolated from respondents with intervention. Therefore, the effects of the intervention might have spilled over to the control group. Still, a possible behavioral effect of the intervention should be much weaker in the control group than for participants with prompts.

The sample of this study consisted of 423 households that were recruited in 3 adjoining areas of peri-urban Cochabamba. The sampling procedure is presented in Figure 3. As in Study 1, a random route approach (e.g., Hoffmeyer-Zlotnik 2003) was used for sampling. Since almost half of the households of the three areas were recruited, no effects of spatial clustering are expected. The refusal rate was 35.2%. However, from households that refused to participate in the survey, data were gathered with an extremely short questionnaire resulting in complete data for the investigation presented. Data were gathered in three panel waves in intervals of a month. Local students performed the interviews with the person responsible for the drinking water of each household. In addition to the panel survey, 170 randomly selected households were monitored twice a week. However, unlike in Study 1, the monitoring data of the current study were gathered in face-to-face interviews (in contrast to self-administered diaries), due to high rates of illiteracy. The interviews comprised 10 closed and 3 open items to ask about behavior, attitudes, norms, communication, and events. The interventions targeted all households in the sample. However, only 192 cases actually received a prompt, while 203 households remained without one.

Figure 3. Flow chart of the sampling procedure in Study 2. Note: *149 households (35.2%) refused to participate in the extensive version of the interview but completed a shorter version, containing only the most important items.

As in Study 1, the panel surveys comprised many more items, but only behavior and the two attitude components are discussed here. *Behavior* was measured by self-report. Participants were asked whether they had applied SODIS during the past month (*yes* = 1 or *no* = 0). The item for the *affective attitude* asked if and how much the participant felt like using SODIS (four response options ranging from 0 to 1). The *instrumental attitude* was measured by asking how much respondents felt it was worth purifying the water with SODIS (seven response options ranging from -1 to 1).

The effects of the monitoring and the intervention on these three constructs were investigated with the data of Panels 2 and 3 only, since sample sizes for the attitudes were minimal in Panel 1. Taking into account that the assumptions for a two-factorial univariate analysis of variance with repeated measures were not met, Wilcoxon Signed Ranks Test (for within-group effects)

Table 3. Means, Standard Deviations, and Number of Cases in the Panel for the Groups of Study 2

		Behavior			Affective Attitude		Instrumental Attitude	
		Monitoring Panel			Monitoring Panel		Monitoring Panel	
Control Panel 1	Count	0	7	<i>M</i>	0.500	0.408	0.585	0.702
	%	0	4.2	<i>SD</i>	0.240	0.280	0.170	0.398
	Cases	10	166	Cases	2	9	4	10
Control Panel 2	Count	6	12	<i>M</i>	0.594	0.561	0.594	0.787
	%	42.9	8.5	<i>SD</i>	0.150	0.251	0.279	0.183
	Cases	14	142	Cases	9	31	9	31
Control Panel 3	Count	13	38	<i>M</i>	0.592	0.646	0.644	0.611
	%	86.7	29.7	<i>SD</i>	0.201	0.165	0.214	0.288
	Cases	15	128	Cases	13	58	13	58
Intervention Panel 1	Count	10	3	<i>M</i>	0.576	0.670	0.543	0.750
	%	10.8	5.7	<i>SD</i>	0.302	0.000	0.460	0.320
	Cases	93	53	Cases	47	4	48	4
Intervention Panel 2	Count	80	21	<i>M</i>	0.702	0.565	0.775	0.692
	%	63.5	32.8	<i>SD</i>	0.257	0.284	0.207	0.281
	Cases	126	64	Cases	109	29	109	29
Intervention Panel 3	Count	119	34	<i>M</i>	0.724	0.716	0.763	0.751
	%	93.7	58.6	<i>SD</i>	0.182	0.235	0.200	0.250
	Cases	127	58	Cases	127	28	127	28

Note. For the dichotomous behavior, the number (Count) and percentage (%) of people who did SODIS is listed.

and Mann-Whitney *U* Tests (for between-group comparisons) were calculated for the attitude variables. Chi-square tests were applied for the dichotomous behavior measure. The effect sizes are defined as in Study 1.

Results

The panel survey resulted in 395 valid cases (93.4% of paneled sample) of which 142 (83.5% of the monitored sample) additionally provided a sufficiently complete (less than three sequent missing values) monitoring time series. About 127 households of the monitored and 65 of the paneled respondents implemented prompts, whereas 15 participants of the monitored and 188 paneled households remained without intervention. In Table 3, descriptive statistics are presented for the three variables—behavior, affective, and instrumental attitude—divided into the three panel waves and the four investigated groups: with versus without intervention and with versus without monitoring. The means are further plotted in Figure 4.

Figure 4. Percentages (behavior) and means (attitudes) of three variables at the three measurement times for the groups of Study 2.

As in Study 1, a strong intervention effect on behavior was observable in Figure 4, which was even more pronounced in the monitored group. In fact, almost the same percentage of monitored cases without intervention performed the behavior at the final panel survey, as monitored cases in the intervention condition. Effects on the attitudes were less clear. Affective attitude increased for all groups, except for cases with intervention and without monitoring, who first displayed a decrease in affective attitude, followed by a steep increase. The effects were weak, with exception of the cases with no intervention and no monitoring, for which a medium effect was found. Instrumental attitude increased for monitored participants, whereas for cases with neither monitoring nor intervention, instrumental attitude first increased and then declined. Again, effects were weak. Finally, in the intervention group without monitoring, instrumental attitude remained approximately the same over time.

In contrast to Study 1, statistical analyses (Table 4) revealed strong effects of monitoring on behavior. In the monitored group, a much higher percentage of respondents performed the behavior than in the unmonitored group. The effect was so strong that intervention effects were not detectable for monitored subjects. However, monitoring only had a minor impact on attitudes. Regarding the intervention, prompts significantly increased the number of behavior performers. However, an increase in behavior performers over time was also observed in the control group, although to a much

Table 4. Results of the Tests of Effects With Chi-Square Test, Wilcoxon Signed Ranks Test, and Mann-Whitney U Test in Study 2

Panel	Group	Effect	Test	Behavior		Affective attitude			Instrumental attitude		
				χ^2	p	U	Z	p	U	Z	p
2	I	G	M	16.043*	0.000	1'147*		.008	1'359		.188
3	I	G	M	34.252*	0.000	1'758		.909	1'771		.970
2-3	I	T	I	17.576*	0.000		-1.843	.065		-1.059	.290
2-3	I	T × G	M × I	0.057	0.811	876*		.005	1'085		.151
2	C	G	M	—*	0.002	136		.885	86*		.043
3	C	G	M	18.998*	0.000	346		.521	366		.836
2-3	C	T	(I)	19.124*	0.000		-1.558	.119		-1.968*	.049
2-3	C	T × G	M × (I)	—	0.195	91		.150	94		.252
2	M	G	I	2.264	0.132	338		.070	311*		.039
3	M	G	I	—	0.285	583*		.027	591*		.043
2-3	M	T × G	I	—	0.556	439		.548	476		.871
2	P	G	I	19.462*	0.000	435		.784	380		.234
3	P	G	I	14.083*	0.000	644		.053	589*		.017
2-3	P	T × G	I	2.150	0.143	243		.083	204*		.014

Note. If no χ^2 values are presented, the assumption for Pearson χ^2 (expected frequencies in all cells >5) was violated and thus, the p values represent exact significances, calculated by Fisher-Yates Tests.

The first column indicates the measurement points considered. The second column shows the groups. I represents cases with intervention and C (control group) represents cases without intervention. In both groups, cases with monitoring are contrasted with cases without monitoring. M stands for cases with monitoring and P (only panel) for cases without monitoring. In these two groups, cases with intervention are contrasted with cases without intervention. Effect indicates what effects are tested: effects between groups (G), over time (T), or interaction effects of group with time (T × G). The fourth column states what is tested in terms of this article: the effect of the monitoring (M), the intervention (I) or the interaction of monitoring and intervention (M × I). Since the control group did not receive an intervention, the I in parentheses. Results indicating significant effects are marked with an asterisk (*).

lesser extent than in the intervention group (see Figure 4). This might be explained by participants in the control group being located in the same area as those in the intervention group. Although these households had not received prompts, the increasing number of neighbors performing the behavior may have created a descriptive norm to use SODIS and possibly served as a social memory aid (Schaefer and Laing 2000), which prevented subjects forgetting the behavior. The intervention had a significant impact on the affective attitude of the monitored group in the third panel and on the instrumental attitude in the second and the third panel wave. The intervention also affected instrumental attitudes in the paneled sample.

In this study, the monitoring survey led to a rather strong reactivity: the behavior performance was much higher for the monitored than the unmonitored group. However, only few effects of the monitoring survey on the attitude variables were observed.

Study 3: Promoting Solar Water Disinfection (SODIS) in Chuquisaca, Bolivia

Campaign and Method

As in the second study, the campaign during which data were gathered in the third study promoted the use of SODIS. The campaign was conducted in the rural department of Chuquisaca in Bolivia and lasted from May 2007 to March 2008. Again, the intervention comprised the distribution of memory aids, which in this case were cuboids made of cardboard of about 15 cm $\times$ 15 cm $\times$ 30 cm in size with pictures and information on SODIS on each side. These memory aids were distributed by promoters who also explained how to do SODIS. Furthermore, radio advertisements promoted the use of SODIS during the entire campaign. These ads could also be received by the persons in the control group. Nevertheless, the memory aids should have had an additional effect on these radio ads.

The sampling procedure is presented in Figure 5. Within the intervention area (two regions in the district of Zudáñez), nine villages² were selected that were suitable for the investigation. Due to the small size of the villages, no actual sampling took place. All households for which the person responsible for the provision of drinking water could be reached entered the sample (refusal rate = 36.0%). Of the 262 households, 45 were randomly assigned to receive a monitoring survey besides the panel survey. According to the campaign plan, all households in the sample should have received an intervention during two time periods. However, only 64 participants

Figure 5. Flow chart of the sampling procedure in Study 3.

received an intervention in both periods (intervention group), whereas 89 cases did not receive any intervention (control group). The households that received an intervention in only one of the two periods were not used in this investigation. About 25 households participated in the monitoring survey, 19 with intervention and 6 without. All cases were surveyed in 3 panels every 4 months. As in Study 2, the data for the monitoring surveys were gathered by face-to-face interviews. However, the interviewers visited the households only once a week, instead of biweekly. Furthermore, the questionnaire for the monitoring survey was much longer than the one used in the previous studies. It comprised approximately 60 closed and 8 open items. Behavior was assessed on a daily basis between the interviews. Many items referred to communication events that could have happened in varying frequencies. Therefore, the actual number of items varied. As in Studies 1 and 2, behaviors, attitudes, norms, and events were assessed.

In the panel surveys, the behavior and psychological constructs under investigation were assessed with multi-item scales that comprised items asking about doing SODIS and drinking untreated water. *Behavior* was assessed with 2 items asking about the quantity of SODIS-treated and untreated water consumed, respectively (each on a 5-point scale). The scale for *affective attitude* comprised 6 items asking about how much they like SODIS-treated and untreated water, how it tastes, and how good or bad they think it is (each on a 9-point scale). The *instrumental attitude* was assessed with 4 items asking about the health effects of drinking untreated or SODIS-treated water (each on a 9-point scale) and the cost and difficulty of doing SODIS (each on a 5-point scale). On all these scales, a value of -1 represents 100% consumption of untreated water or an attitude completely in favor of untreated water and rejecting SODIS. A value of 0 represents equally strong behaviors or attitudes for SODIS and drinking untreated water, whereas a value of $+1$ means 100% SODIS use or a completely favorable attitude toward SODIS and against drinking untreated water.

As in Study 1, the reactivity of the monitoring survey and the intervention effects were quantified with the General Linear Model. Since the number of cases with monitoring without intervention was too small, only cases with intervention were investigated regarding the reactivity of the monitoring survey. Only cases without monitoring were considered to quantify the intervention effect. The effect sizes are defined as in Studies 1 and 2.

Results

The numbers of cases that entered the analyses are compiled in Table 5. Dropouts were mostly caused by logistic problems. Missing values in the attitude variables, particularly in Panel 1, are due to the fact that persons who did not know SODIS were not asked about any attitudes toward this water treatment. Table 5 compiles the descriptive statistics for the three variables investigated divided into the panel waves and groups as in the previous studies. The means are plotted in Figure 6.

Figure 6 shows surprisingly strong changes over time for the behavior as well as for the attitudes in the control group. Since SODIS did not spread before the campaign, these effects must be due to the radio ads and also spillover effects from the intervention group. Over time, the behavior and both attitude components show strong and medium changes, respectively. However, the monitoring survey seems not to have had any effect.

The observations of Figure 6 are confirmed by the analyses of variance presented in Table 6. All changes over time with exception of the

Table 5. Means, Standard Deviations, and Number of Cases in the Panel for the Groups of Study 3

		Behavior		Affective Attitude		Instrumental Attitude	
		Monitoring	Panel	Monitoring	Panel	Monitoring	Panel
Control Panel 1	M	-0.250	-0.580	-0.031	0.147	0.203	0.379
	SD	0.500	0.423	0.365	0.309	0.219	0.299
	Cases	6	66	4	17	4	17
Control Panel 2	M	0.000	-0.250	-0.042	0.094	0.031	0.488
	SD	0.707	0.581	0.236	0.328	0.044	0.328
	Cases	2	62	2	46	2	46
Control Panel 3	M	0.200	0.212	0.333	0.342	0.563	0.571
	SD	0.512	0.550	0.347	0.316	0.238	0.310
	Cases	5	46	5	44	5	45
Intervention Panel 1	M	-0.289	-0.526	0.071	0.106	0.335	0.524
	SD	0.393	0.357	0.158	0.220	0.246	0.213
	Cases	19	39	14	13	14	13
Intervention Panel 2	M	0.000	0.111	0.162	0.230	0.595	0.632
	SD	0.464	0.565	0.164	0.248	0.201	0.236
	Cases	19	45	19	38	19	38
Intervention Panel 3	M	0.671	0.494	0.467	0.393	0.694	0.589
	SD	0.205	0.363	0.233	0.319	0.179	0.330
	Cases	19	45	19	45	19	44

instrumental attitude in the second period turned out significant, whereas no significant effects of the monitoring survey are observed. Even though the interaction effect of monitoring and intervention for behavior in the first period was almost significant, this effect cannot be found if both periods are analyzed together. However, there was a medium interaction effect in the case of the instrumental attitude: persons with monitoring increased their assessment of the health effects of treated water, whereas persons without monitoring did not change their opinion.

In spite of the much longer interviews performed in this study compared to Study 2, the reactivity of the monitoring survey was negligible. Since the conditions in the two studies were almost the same, the reduction of the reactivity was attributed to the less frequent interviewer visits. It seems that monitoring surveys with face-to-face interviews can be performed without reactivity, if the interviewers visit the households only once a week or less often. Furthermore, this study showed that rather complex questionnaires could be applied even within a monitoring survey that lasts as long as a year.

Figure 6. Means of three variables at the three measurement times for the groups of Study 3. Due to the small number of cases, no mean is presented for the group with monitoring but without intervention in Panel 2.

Discussion

For three different behavior-change campaigns in developing countries, the reactivity of gathering time-series data by monitoring surveys was investigated. Results found that reactivity can be an issue but, nevertheless, monitoring surveys can be implemented with negligible effects on the data. It is possible to gather data even with relatively long face-to-face interviews, but the following settings are problematic: (a) Face-to-face interviews that are performed more often than weekly lead to reminding effects that can increase the behavior performance. Such effects were not observed for self-administered diaries or weekly interviews; (b) In the case of self-administered diaries, negative effects on the attitude were observed for the control group. It appears that persons who do not receive any intervention do not understand the purpose of the frequent data gathering. This, in turn, seems to lead to participation fatigue and resentment. Therefore, traditional control-group designs might not be feasible for monitoring surveys; (c) Face-to-face interviews, in comparison to self-administered diaries, seem to have the inverse effect: they lead to more target behavior favoring attitudes, particularly for the persons with intervention. This effect, however, is small if the interviews are not performed more often than weekly.

Based on the studies presented, it is recommended to combine self-administered and interview-based monitoring surveys with traditional

Table 6. Results of the Tests of Effects With the General Linear Model (GLM) in Study 3

Panel	Group	Effect	Test	Behavior		Affective attitude		Instrumental attitude	
				<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
1–2	<i>I</i>	<i>G</i>	<i>M</i>	0.552	.461	2.464	.130	3.614	.070
1–2	<i>I</i>	<i>T</i>	<i>I</i>	26.096*	.000	6.818*	.015	13.946*	.001
1–2	<i>I</i>	<i>T</i> × <i>G</i>	<i>M</i> × <i>I</i>	3.470	.068	0.563	.460	1.370	.254
1–2	<i>P</i>	<i>T</i> × <i>G</i>	<i>I</i>	4.590*	.035	0.688	.416	0.016	.900
2–3	<i>I</i>	<i>G</i>	<i>M</i>	0.169	.683	0.067	.796	0.596	.443
2–3	<i>I</i>	<i>T</i>	<i>I</i>	32.938*	.000	14.525*	.000	0.263	.610
2–3	<i>I</i>	<i>T</i> × <i>G</i>	<i>M</i> × <i>I</i>	2.453	.122	1.842	.180	1.510	.224
2–3	<i>P</i>	<i>G</i>	<i>I</i>	9.536*	.003	0.047	.829	0.814	.370
2–3	<i>P</i>	<i>T</i> × <i>G</i>	<i>I</i>	0.142	.707	0.189	.666	1.161	.285
1–3	<i>I</i>	<i>T</i> × <i>G</i>	<i>M</i> × <i>I</i>	0.335	.565	1.079	.309	6.143*	.021

Note. The first column indicates the measurement points considered, and the second the groups. *I* represents cases with intervention where cases with monitoring are contrasted with cases without monitoring. *P* (only panel) represents cases without monitoring, where cases with intervention are contrasted with cases without intervention. *Effect* indicates what effects are tested with the GLM: effects between groups (*G*), over time (*T*), or interaction effects of group with time (*T* × *G*). The fourth column states what is tested in terms of this article: the effect of the monitoring (*M*), the intervention (*I*) or the interaction of monitoring and intervention (*M* × *I*). Due to a lack of cases with monitoring in the control group, the monitoring effect for cases without intervention and the intervention effect for cases with monitoring could not be investigated. Results indicating significant effects are marked with an asterisk (*).

panel surveys for gathering time-series data during campaigns in developing countries. Very critical data, particularly on behavior, should be gathered at least on a daily basis without interviewers visiting the households. Even in difficult settings (e.g., with high illiteracy or lack of paper and pencils), simple ways of tracking behavior performance, and maybe even a few other variables, can be implemented. Forms could be designed to be punctured or torn as a means to mark each behavior performance if pencils are scarce, or stones can be filled into jars. If possible, short questionnaires can be used to gather more data, as was done in Study 1. This data gathering should be limited to information that changes rapidly (e.g., behavior or affective attitude) or is difficult to remember (e.g., events or communications).

About every 1–3 weeks, an interviewer should visit the monitored persons. Besides recollecting the aforementioned data, face-to-face interviews allow gathering a larger amount of data on more complex issues. Furthermore, daily data can be discussed. For example, if a communication

was noted by the participant, more details about the information, communicator, or effects can be asked. The face-to-face interviews should also be used to ask open questions, which usually are not well answered in self-administered diaries, and to discuss possible problems (i.e., in the case of missing answers, inconsistencies, or problems mentioned by the participant). Interviewer visits could also be used to gather observational data.

Finally, every monitoring should be combined with a panel survey, which allows collecting an even larger amount of data, and assess constructs that are more difficult to measure. The sample of the panel survey should be much larger than the monitoring sample to allow more elaborated statistical analyses and to test the reactivity of the monitoring. These panel surveys should be conducted at least before and at the end of a campaign or change process. However, for longer investigation periods, a panel survey every few months should be considered. In addition, to reveal long-lasting effects, a survey some months or even years after the end of a campaign should be performed.

It can be concluded that the gathering of time-series data is possible and valuable even under the difficult conditions of campaigns in developing countries. Self-administered diaries minimize the reactivity but are often not feasible due to high levels of illiteracy or other problems. Face-to-face interviews can lead to valid data, if the intervals of visits are not too short (i.e., weekly or less often). Nevertheless, the reactivity of the data gathering should be investigated in any study.

Notes

1. Regarding the investigation presented in this article, no relevant differences were found between the data of the two districts and therefore the intervention group is treated as one unit of analysis.
2. Regarding the investigation presented in this article, no relevant differences were found between the data of the nine villages. Therefore, the intervention group was treated as one unit of analysis.

Acknowledgment

The campaigns during which the data were gathered were initiated by Hans-Joachim Mosler and coordinated by Andrea Tamas (both Eawag). The campaign in Cuba was implemented with the aid of Tamara Caballero Rodríguez and Omar Guzmán Miranda (Universidad de Oriente, Santiago de Cuba). The campaigns in Bolivia were implemented with the aid of the SODIS Foundation (Fundación SODIS), Cochabamba, the local NGO Obispo Anaya, and Carmen Camargo. The authors thank Silvie Kraemer and Hans-Joachim Mosler for helpful comments and discussion.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

This research was supported by Swiss National Centre of Competence in Research (NCCR) North–South: Research Partnerships for Mitigating Syndromes of Global Change, co-funded by the Swiss National Science Foundation (SNF), the Swiss Agency for Development and Cooperation (SDC), and the UBS Optimus Foundation, Switzerland.

References

- Bamberg, S. 2002. Effects of implementation intentions on the actual performance of new environmentally friendly behaviors—Results of two field experiments. *Journal of Environmental Psychology* 22:399-411.
- Bamberg, S., and P. Schmidt. 2001. Theory-driven subgroup-specific evaluation of an intervention to reduce private car use. *Journal of Applied Social Psychology* 31:1300-29.
- Berney, M., H. U. Weilenmann, A. Simonetti, and T. Egli. 2006. Efficacy of solar disinfection of *Escherichia coli*, *Shigella flexneri*, *Salmonella Typhimurium* and *Vibrio cholerae*. *Journal of Applied Microbiology* 101:828-36.
- Binder, C., and H.-J. Mosler. 2007. Waste-resource flows of short-lived goods in Santiago de Cuba. *Resources, Conservation and Recycling* 51:265-83.
- Cialdini, R. B., R. R. Reno, and C. A. Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58:1015-26.
- Hoffmeyer-Zlotnik, J. H. P. 2003. New sampling designs and the quality of data. *Advances in methodology and Statistics/Metodološki zvezki* 19:205-17.
- Hoppe, M. J., M. R. Gillmore, D. L. Valdez, D. Civic, J. Hartway, and D. M. Morrison. 2000. The relative costs and benefits of telephone interviews versus self-administered diaries for daily data collection. *Evaluation Review* 24:102-16.
- Heri, S., and H.-J. Mosler. 2008. Factors affecting the diffusion of solar water disinfection: A field study in Bolivia. *Health Education & Behavior* 35:541-60.
- Joyce, T. M., K. G. McGuigan, M. Elmore-Meegan, and R. M. Conroy. 1996. Inactivation of fecal bacteria in drinking water by solar heating. *Applied Environmental Microbiology* 62:399-402.
- Kubey, R., R. Larson, and M. Csikszentmihalyi. 1996. Experience sampling method applications to communication research questions. *Journal of Communications* 46:99-120.

- Morrison, D. M., B. C. Leigh, and M. R. Gillmore. 1999. Daily data collection: A comparison of three methods. *Journal of Sex Research* 36:76-81.
- Moser, S., and H.-J. Mosler. 2008. Differences in influence patterns between groups predicting the adoption of a solar disinfection technology for drinking water in Bolivia. *Social Science & Medicine* 67:497-504.
- Mosler, H.-J., S. Drescher, C. Zurbrügg, T. Caballero Rodriguez, and O. Guzman Miranda. 2006. Formulating waste management strategies based on waste management practices of households in Santiago de Cuba, Cuba. *Habitat International* 30:849-62.
- Mosler, H.-J., A. Tamas, R. Tobias, T. Caballero Rodriguez, and O. Guzman Miranda. 2008. Deriving interventions on the basis of factors influencing behavioral intentions for waste recycling, composting and reuse in Cuba. *Environment and Behavior* 40:522-44.
- Schaefer, E. G., and M. L. Laing. 2000. 'Please remind me . . .': The role of others in prospective remembering. *Applied Cognitive Psychology* 14:S99-S114.
- Schmitz, B., P. Stanat, F. Sang, and K. Tasche. 1996. Reactive effects of a survey on the television viewing behavior of a telemetric television audience panel. *Evaluation Review* 20:204-29.
- Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18:429-34.
- Tobias, R. 2009. Changing behavior by memory aids: A social psychological model of prospective memory and habit development tested with dynamic field data. *Psychological Review* 116:408-38.
- Tobias, R., A. Bruegger, and H.-J. Mosler. 2009. Developing strategies for waste reduction by means of tailored interventions in Santiago de Cuba. *Environment & Behavior* 41:836-65.
- Wegelin, M., S. Canonica, K. Mechsner, T. Fleischmann, F. Pesaro, and A. Metzler. 1994. Solar water disinfection: Scope of the process and analysis of radiation experiments. *Journal of Water Supply: Research and Technology—Aqua* 43:154-69.

Bios

Robert Tobias, PhD in psychology and MS in civil engineering, is senior scientist at the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) in the Department of System Analysis, Integrated Assessment, and Modeling (SIAM), where his primary task is the modeling of behavior-change campaigns. His research interests comprise behavior change, communication and persuasion, prospective memory and habits, social norms and collective actions, social networks and diffusion of innovations, agent-based simulation, and measurement of psychological constructs and their dynamics.

Jennifer Inauen, MSc in Psychology, is a PhD student at the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) in the Department of System Analysis, Integrated Assessment, and Modeling (SIAM). Her current research is concerned with the effectiveness and mechanisms of behavior-change interventions to enhance the habitual use of arsenic-safe water sources in Bangladesh.