PDAs in Socioeconomic Surveys: instrument bias, surveyor bias or both?

Javier Escobal & Sara Benites

GRADE, Lima, Peru

The authors are, respectively, Senior Researcher and Assistant Researcher at the Group for the Analysis of Development (GRADE). Javier Escobal is Principal Investigator of Young Lives in Peru. Correspondence to: Javier Escobal, GRADE, Av. Almirante Grau 915 Barranco, Lima 4, Peru. Tel.: +51 (1) 247 9988; Fax: +51 (1) 247 1854; Email: jescobal@grade.org.pe.

Abstract:

It is very likely that electronic means of data collection will become a standard for socioeconomic surveys in the near future. As surveys move from paper form to personal digital assistants (PDAs) or tablets there is a need to evaluate if such a shift will affect the quality of the data. To explore the potential biases that data collection through PDAs may generate, we use the third wave of Young Lives-Peru to randomly assign survey respondents between PDA-based questionnaires and paper-based questionnaires. The study shows that data collected through PDAs has a higher level of agreement when compared to paper-based responses. However, there are a number of situations where paper-based questionnaires do not easily translate to PDA-based questionnaires. In addition, the study highlights the need to improve training as different surveyor's characteristics (age, gender, education, etc.) may also be correlated with some observed differences between PDA and paper-based questionnaires.

Keywords: PDA, survey forms, report bias, Peru

Introduction

In this paper we explore the extent in which PDAs may affect the quality of the data collected in surveys. For this we use the Peruvian Young Lives study. Young Lives is an international study carried out in four countries (Ethiopia, Vietnam, India and Peru) whose objective is to improve the understanding of the causes and consequences of childhood poverty and to examine how economic and social policies affect children's well-being. In Peru, Young Lives has been tracking about 2000 children from a younger cohort aged between 6 months and 18 months in 2002, at the time of the enrolment. It also has been tracking an older cohort of about 700 children aged between 7.5 years old and 8.5 years old in 2002. A second visit was carried out from late 2006 to early 2007; a third visit was executed in late 2009 and subsequent visits are scheduled for 2013 and 2016.

Recent advances in technology related to the management of surveys are an important progress in terms of reducing time for data availability and improving their quality. The range of technologies available is large; from traditional paper-based questionnaires and surveys via telephone or computer, to the self-administered surveys that use the internet.[1] However, not all types of technologies are equally reliable in developing countries due to physical, geographic, economic or even social constraints.

While the paper and pencil interview (PAPI) allows face-to-face interviews to be more effective and can withstand questionnaires with a high level of complexity, inconsistencies, data entry errors and blank responses are an important drawback. On the other hand, in computer-assisted telephone interviews (CATI), long questionnaires are difficult to implement. Further, limited coverage of land-line phones and low reliability of cell phones in many rural areas affect the sampling design and the quality of data obtained through this method, generating lower response rates and a higher response bias. Computer-assisted self-interviewing (CASI) will also be affected by a lower response rate and a higher level of

response bias, as it favours better educated and literate population and will generate under-coverage of those people with no access to internet access.

Many are the benefits that are typically associated with the use of the PDAs, which fall within the CAPI group (Computer-assisted personal interview). Higher accuracy, faster data retrieval and reduced costs are three potential benefits (Stolworthy, 2003). As for higher accuracy, one should expect lower number of inconsistencies and missing values with PDA-based questionnaires. For example, the software that runs the survey can assure that all questions in a certain section are filled before proceeding to the following section. In addition, certain data like the time the interview starts or the GPS coordinates of the dwelling in which the survey is administrated can be automatically recorded eliminating errors or fraud. Finally, if faster data retrieval is obtained, double data entry is no longer necessary.

According to Lane, Heddle, Arnold and Walker (2006) the benefits are clear as:

> …handheld computers can be programmed to provide determinate responses, date stamped to document times of data entry, restrict times of data entry, prevent retroactive data entry, limit 'look back' to previous data, prevent omissions of data entry, and can save considerable time and labour incurred in data handling. (p. 8)

A number of authors have expressed positive views around the use of PDAs to collect data. (i.e. Shapiro, Bessette, Baumlin, Ragin, & Richardson (2004); VanDenKerkhof, Goldstein, Blaine, & Rimmer (2005)). Becker, Gries, Martin and Bernhard (2010) discuss the experience of using PDAs to gather patient data and highlight their benefits in terms of customizability. Similarly, Lu, Xiao, Sears and Jacko (2005) highlight its benefits although raise the concern that there is a need to improve PDA hardware, software and institutional support to maximize their benefits.

Disadvantages of using a PDA-based questionnaire do exist: they hinder readability due to the small screen size, decrease the flexibility to skip questions, or do not allow certain answers which are not allowed by default. Galliher et al. (2008) recognise that they may be

other reasons why PDAs may not perform as paper form questionnaires, like missing data due to programming errors or other technical difficulties, or even the complete loss of the data due to PDA malfunctioning, loss or theft.

Further, although PDA-based questionnaires may assure that missing data is low, one needs to take careful consideration if this lower missing data report is correlated or not with enhanced data quality. It is clear that forcing the surveyors to mark data does not necessarily imply an improvement on its quality.

A number of studies have formally evaluated the differences between PDAs and paper forms as data recording instruments. Dale and Hagen (2007) review nine studies and find that, despite some technical problems, PDAs outperform pen and paper when collecting patient data because PDAs lead to improved protocol compliance. Lane et al. (2006) review nine studies (only two of these coincide with the review made by Dale & Hagen (2007)), and show that missing data is substantially lower in PDA-based questionnaires.

E. Juniper, Langlands and B. Juniper (2009) show in a small sample context that differences between paper and electronic forms were significant when analysing a group of individuals that were randomized between paper and electronic versions of the same set of questionnaires associated with asthma and rhinitis. The reason for this bias was reported as unclear. Fletcher, Erickson, Toomey and Wagenaar (2003) show that although the agreement between methods is very high, it has decreased as forms have increased in length and difficulty.

Bernabe-Ortiz et al. (2008) compare data concerning sexual behaviour collected with paper forms to data collected with PDA-based forms in Peru. In a first study, each questionnaire was self-administrated in both formats. Although the paper self-applied questionnaire was closed and put into a locked voting bag, there was no guarantee that the same level of confidentiality was obtained using the PDA. In a second study reported by the

same authors, both PDA and paper forms were administrated by the surveyors to 50% of the sample. The agreement is higher in the second study which may suggest that in the first one the usage of PDA could be associated with a higher likelihood of perceiving a confidentiality breach.

Time saving may also be an important issue, as the total time of administration and required manual transcription onto a computer database of the paper-based questionnaires is higher. Rivera et al. (2008) report for a randomized albeit small sample study, that the total data gathering time, inclusive of transcription, was extensively less for the PDA with a significant improvement on the data integrity.

The efficacy of the PDA may also depend on the type of data been collected. Jaspan et al. (2007) look at the benefits of the PDA for self-administrated surveys on sexual behaviour. Although the research staff in this evaluation continuously emphasised that both data collection methods were equally confidential and anonymous, a considerably smaller proportion of participants reported ever having had intercourse when answering a PDA-based questionnaire, in comparison to a paper-based questionnaire that after finished was folded and introduced into a sealed box.

On the other hand, despite the fact that Computer-assisted personal interviews (CAPI) using personal computers (laptops) share some advantages with the PDAs, there have some disadvantages, especially in developing countries, including: higher hardware costs, lower battery duration and higher visibility. In that sense, while PDAs pose extra challenges to the surveyors, who could show less familiarity with the advanced technology and need higher levels of education to deal with it, laptops show less practicality, are more difficult to operate in some circumstances and are more prone to theft because their visibility.

Considering that most, if not all, of the PDA assessments done are associated to their use in clinical research; we believe that it is useful to extend such evaluations to

socioeconomic surveys. In addition, although everybody will agree that the quality of a survey is strongly associated with the quality of training and experience of the surveyors, the literature has not addressed how the instrument used to collecting data may be affected by the characteristics of the surveyors. To address these issues the paper is divided in three sections including this introduction. In section 2 we present the general methodology used to evaluate the potential biases that data collection through PDAs may generate. This section considers that three types of biases may appear: instrument bias, surveyor bias and small sample bias. Next, in section 3 we present the results of randomizing PDA-based and paper-based questionnaires for both the younger and the older cohort of the Peruvian Young Lives study. Finally, section 4 concludes with a summary of our findings and key recommendations that can be derived from the analysis.

Methodology: Assessing the impact of using PDAs

Suppose that the recording of *y*, an outcome of interest, is affected depending on whether it is registered through a PDA-based questionnaire or a paper-based questionnaire. We observe outcome y_{1i} for a respondent i whose information has been recorded through a PDA-based questionnaire and we observe y_{0i} for a respondent i whose information has been recorded through a paper-based questionnaire. Denote the dummy variable T_i, which has a value 1 if the information has been recorded through a PDA-based questionnaire and 0 if not. Formally, for each respondent we have:

$$y_i = y_{0i} + (y_{1i} - y_{0i}) \cdot T_i \tag{1}$$

First, let define the usage of a PDA or the nature of the intervention, as the "treatment", and the one who did actually experience that intervention as the "treated". In contrast, the "control" group would be the one that did not participate in the intervention.

Since for a particular respondent either y_1 or y_0 is observed, the average effect of the treatment on the treated (ATT) refers to the average response to treatment for a sample of individuals that were assigned to a PDA. Further, if we measure the unconditional expected difference between outcomes we obtain the average treatment effect (ATE), which is the average response to treatment for a random sample from the population. In this case ATE is the average effect of using a PDA when compared to those whose information was captured using a paper-based questionnaire.

Unfortunately, those whose data was captured using a paper-based questionnaire will not necessarily be comparable to those whose data was obtained using a PDA because they may differ in terms of age, ethnicity, socioeconomic status, etc. If this is the case, the difference between the average outcome of those whose data was obtained using a PDA and the average outcome of those whose data was obtained by a paper-based questionnaire will be a biased estimation of the effect of using PDAs.

There are two ways one can control for observable attributes and make both samples comparable. One possible way is to estimate a regression of the form:

$$y_i = \alpha + \delta T_i + X_i \beta + u_i \tag{2}$$

Here the coefficient δ is the estimate of the treatment effect or the effect of using a PDA. In order to obtain a consistent estimator of δ, the key assumption is that the treatment (T_i) is not correlated with the unobserved determinants of the outcome (u_i). If the treatment variable, which indicates which respondents were interviewed with a PDA and which were interviewed using a paper-based questionnaire, is correlated with u_i we will need additional information to obtain a consistent estimator of the effect of using a PDA. This information comes through an extra variable (called instrumental variable or IV), that should be

uncorrelated with u_i but should be correlated with the assignment rule. The IV is used to identify exogenous variation in the outcome variable.

Another possible approach is to limit our comparison to respondents who are comparable, in terms of some key characteristics. These type of estimators are called matching estimators. The advantage of this second approach is that we do not have to make any assumption regarding the functional form used to estimate (in this case a linear functional form). However, the matching technique will give a consistent estimate of the average effect of using a PDA only under the assumption that selection process depends on observables and that the procedure used to match respondents assigned to PDAs and respondents assigned to paper-based questionnaires is the appropriate one.

Adjusting the methodology to incorporate specificities of the Young Lives field work

In order to evaluate the effect of using the PDAs, a set of Young Lives questionnaires that capture relevant socioeconomic information were transcribed into PDAs following exactly the question order and the same wording for each question and for all available options and categories. Having both formats, half of the sample was randomly assigned to PDA-based questionnaires and the other half was assigned to paper-based questionnaires. In order to perform the randomization we used as an instrument a variable that indicates whether the number assigned to the child (child-id) is odd or even. Children with even child-id numbers were assigned to PDAs and children with odd child-id numbers were assigned to paper-based questionnaires.

To confirm that the assignment was random it is important to understand how these child-ids where assigned. In Peru, the sampled areas for the Young Lives project were chosen using a multi-stage, cluster-stratified, random sampling approach.[2] The sample was divided into 20 clusters. Within each cluster, when the surveyors entered into an area to seek children

of the specified age (between 6 and 18 months of age for the younger cohort and between 7.7 and 8.5 years for the older cohort), the child-id numbers were assigned following a first-come first-serve basis, the one that the surveyors managed to secure the interviews. The surveyors started in the centre of the sampled area and then swept it increasing the radius, as they confirmed that there were no additional children of the chosen ages available. Because of the way the id codes were assigned, odd and even ids show a similar pattern of geographic dispersion.

If we believe that the assignment is indeed random, there should be no correlation among the assignment between PDA and paper-based questionnaires and the error term in (2). If this is the case, for any outcome, the difference of the averages between those interviewed using PDAs and those interviewed using paper forms should give us a measure of the impact of using PDAs.

Although the majority of the questionnaire assignment followed the proposed rule, we do have a small percentage of questionnaires showing non-compliance. For both cohorts, 3% of the questionnaires were incorrectly assigned. Just a few cases (five) correspond to questionnaires that were supposed to be done in paper but because of logistics a paper form was not available when the surveyor was arranging the day and time of the interview and the respondent insisted to be interviewed at that moment. All the other cases correspond to an assignment of interviews that were supposed to be done using a PDA-based questionnaire but were finally done using a paper-based questionnaire. Most of these cases correspond to respondents who were interviewed at the very last stage of the gathering period and were part of those who had migrated away from the original sites to isolated or dangerous areas. For these cases, the survey management team decided to assign more experienced albeit less technologically savvy surveyors to these interviews.

Although small, this non-compliance when comparing both sub-samples raises a more important point regarding the fact that even if questionnaires were, in most cases, randomly assigned between PDA-based and paper-based questionnaires, surveyors were not randomly assigned to interviews. Thus, differences between PDA-based and paper-based questionnaires may also capture differences in surveyor's characteristics. Even if we were able to randomize most of the questionnaire forms, surveyor's characteristics do show systematic differences between the two methods, which may affect the interviews beyond the type of questionnaire used. In particular, those assigned to a PDA are more likely to be men, younger, more educated and less likely to have children. Although all surveyors were trained in the usage of the PDA, a subset of them (all women) did not feel at ease with PDAs. Despite their extensive experience in the management of surveys and field work, the team decided to let them conduct surveys using only paper-based questionnaires. This, of course, may potential bias the evaluation and needs to be taken into consideration.

In addition, we developed two measures of performance that could help us to distinguish between more skilled and less skilled surveyors. The first one is a measure based on the errors spotted in the implementation of the Peabody Picture Vocabulary Test (PPVT). The PPVT is a norm-referenced test used to assess child receptive vocabulary (i.e., it measures listening comprehension of spoken words), a key indicator of cognitive development, and it was done in paper form for the entire sample. Considering that there is a very strict protocol that must be followed to administer the test, it is easy to spot and register any error made. The second measure is a qualitative assessment of all of the surveyors after a detailed scrutiny of their questionnaires. Surveyors were divided into three groups, according to their relative performance, the number of errors made in the survey administration and the number of inconsistencies found when reviewing the questionnaires.

Although we do not find significant differences in PPVT performance for surveyors assigned to either type of questionnaire, there is some evidence that a larger share of surveyors that were later evaluated as less skilful were, in fact, assigned to paper-based questionnaires.

As we have mentioned, even if we were facing a complete randomization we could observe differences in respondent characteristics among the two methods. To avoid any potential endogeneity we have compared only first round characteristics, that is, characteristics of the respondents when they were enrolled into the Young Lives Survey. These characteristics are: area of residence (urban or rural), altitude of the town where the child lives (as a measure of remoteness), region of residence (coast, highland or amazon), household size, a measure of wealth, household total expenditure, gender of the household head, mother's characteristics (age, education and ethnicity) and child's characteristics (gender and nutritional status as measured by height for age). It is observed that for all of these variables, except gender of the household head for the older cohort, we found no significant differences in respondent characteristics between the two types of questionnaires.

Still, if for the older cohort there are differences in the gender of the household head, such differences may affect many other variables reported by these respondents. Given this, we may need to adjust our estimates for this potential small sample bias.

Disentangling the PDA effect, the surveyor effect and the small sample bias effect

So far we have mentioned that through a regression technique that controls for observable characteristics, or through limiting our comparison to respondents who are comparable in term of some key characteristics, we can estimate the impact of PDAs in the quality of the data. We have also mention that there are three distinct effects that may be affecting the treatment: (1) the effect of using a PDA as opposed to a paper-based questionnaire; (2) a surveyor effect; and, (3) a small sample bias that may exist in a context of respondents being

randomized between PDA and paper-based questionnaires.

If we expand equation (2) to incorporate a set of variables to control for surveyor's characteristics (Z) and the interaction of the treatment effect with Z and first round characteristics (X), we obtain the following:

$$y = \alpha + \delta T + \sum \beta_i X_i + \sum \gamma_i X_i T + \sum \lambda_i Z_i + \sum \varphi_i Z_i T + u \quad (3)$$

If the assignment between those interviewed using PDAs and those interviewed using paper is based on a randomization and therefore is uncorrelated with the error term, Ordinary Least Squares (OLS) will provide unbiased estimates. However, if the treatment assignment is not completely random, the equation can be estimated using an appropriate instrument. In our case such instrument exists and is simply the variable that indicates whether the child-id is odd or even. This variable is highly correlated with T and should not be correlated with the error term.

We can use (3) to test each of the effects discussed above:

- PDA effect is not present : $\delta=0$ and $\gamma = 0$ and $\varphi = 0$
- Surveyor effect is not present : $\lambda = 0$ and $\varphi= 0$
- Small sample bias is not present : $\beta = 0$

In the next section we use this framework in order to estimate the average treatment effect of using PDAs, first under different matching estimators and then with appropriate instruments. Further, we estimate and discuss whether the differences found can be attributed to a small sample bias, to a surveyor effect or if they constitute effects that can be attributed to the usage of PDAs instead of paper forms.

Main results

First, we estimate the effect of using a PDA-based questionnaire under different matching estimators. In all cases we first estimated the propensity score of being treated, i.e. the probability of a respondent being interviewed with a PDA-based questionnaire. In order to implement the estimation, we included as an instrument whether the child-id is odd or even as well as surveyors' characteristics and first round characteristics. In all cases the estimated probabilities were forced to lie within a common support. This excludes the treated individuals whose probabilities of participation are higher than the highest probabilities of the comparison group and the control individuals whose probabilities are lower than the lowest probabilities of the treated group. That is, we confirmed that the range of the estimated probabilities of been interviewed using a PDAs or a paper form was similar for both treated and potential controls.

We explored three matching techniques: a) nearest neighbour, b) stratified matching and c) kernel matching. As is well known (Baser, 2006), all of these techniques implicitly assume a trade-off between bias and efficiency. Kernel matching incorporates a larger number of controls for each treated observation (with appropriate weights) which generates more efficiency but with eventually some bias, in comparison to the nearest neighbour or stratified matching. On the other hand, nearest neighbour matching minimizes the bias by comparing each treated observation with the most similar control. However, by using a reduced number of observations it may potentially produces an estimation of the effect of using PDAs that is less efficient than the one obtained through kernel or stratified matching. Finally, the stratified matching compares each treated observation with those that are part of the same strata, which generates an estimator that lies among the other two in terms of the trade-off between biasness and efficiency. Since no matching estimator is consistently superior to the others, showing all of them allow us to verify how robust the estimated effects are.

We estimate the effect of using PDAs instead of paper-based questionnaires for a sample of 60 outcomes for the younger cohort and a sample of 81 outcomes for the older cohort. Table 1 and Table 2 show only those outcomes that where statistically significant for at least one method for each cohort, respectively. As can be seen in Table 1, for the younger cohort the mean comparison test between those subjects interviewed with a PDA and those interviewed using a paper-based questionnaire shows nine outcomes that are significantly different. These results stand out when we explore the three alternative matching techniques, were we found between 12 and 14 significant outcomes. Taking into consideration that for a significance level of 90% we could expect that on average 10% of the outcomes could be significant just because of chance these results are higher, as between 20% and 23% of the outcomes show a significant difference.

[Table 1 here]

[Table 2 here]

Similar results are obtained from the older cohort where between 14% and 16% of the outcomes show some significant results, a rate that is considerable larger than a significance level of 5% or even 10%. Thus, the results may suggest that although a large percentage of the outcomes are similar we do have a group of questions that differs when using alternative ways of capturing the same information.

The questions that appear to be significant between at least two of the three matching techniques can be grouped in four categories: a) questions related to the child's time use and child work; b) socio-economic shocks experienced by any member of the household; c) degree of risk aversion; and, d) support networks and political capital. In the case of the older cohort there are additional questions related to the access to key programmes (like JUNTOS, the Peruvian conditional cash transfer programme, and SIS, a universal health programme for children).

Before exploring why these outcomes differ between the two collection methods, we present in Table 3 and Table 4 our estimates of the average effect of PDAs using the regression framework presented in equation (3), for the younger and older cohorts, respectively. Table 3 shows that for the younger cohort only 8% of the outcomes are significantly different among methods, a number reasonably low. The outcomes showing highly significant differences are related to the child's time use and socio-economic shocks experienced by any member of the household, a similar result to the one reported using matching estimators.

[Table 3 here]

Table 4 shows that for the older cohort the number of significant differences in outcomes between methods is much larger than the one observed for the younger cohort. Here 20% of the outcomes are significantly different among methods, a number that is much larger than a significance level of 5% or 10%. For the older cohort the outcomes that show significant differences can be grouped again in three categories: a) questions related to the child's time use and child work; b) socio-economic shocks experienced by any member of the household; and, c) support networks and political capital. In this case the variable related to risk attitudes is no longer significant.

[Table 4 here]

If we want to disentangle the transmission channels that are generating the differences in these outcomes we need to test whether these are due to differences in surveyor characteristics, small sample bias or there are truly differences that can be linked to the different way the questionnaire was administrated. Tables 5 and 6 show, precisely, the results from the interaction model only for those outcomes that we found to be significant in Tables 3 and 4.

In the case of the younger cohort, the results indicate that both the variables related to the child's time use and socio-economic shocks experienced by any member of the household show a significant PDA effect and a significant surveyor effect. The variable related to missing schooling shows some evidence of small sample bias in addition to a significant PDA effect. Finally, the asset indicator and whether the children in the household were affected by food problems show no significant PDA effect.

[Table 5 here]

In the case of the older cohort the results related to the child's time use and socio-economic shocks experienced by any member of the household show a similar pattern as the one reported for the younger cohort. We also find both a significant PDA effect and a significant surveyor effect in the prevalence of child work and the request for help in episodes of child abuse or family violence. A number of variables related to access to services (the National Identity Card or the SIS Health System) show significant effects for the three transmission channels. The rest of the variables, with the exception of internet usage, taken part in a protest march and smoke habits, show a significant surveyor bias.

[Table 6 here]

It is interesting to note that if we estimate our treatment model using an instrumental variable approach and test for the advantage of considering our treatment variable as endogenous (using a Wu-Hausman F test or a Durbin-Wu-Hausman Chi-squared test), we cannot reject the null hypothesis, implying that an ordinary least squares model that estimates the treatment effect is as good as the instrumental variable model. Similarly, in all the instrumental variable models tested that included surveyor fixed effects, the Davidson-MacKinnon test of exogeneity shows that we cannot reject the null hypothesis, implying that our randomization has not been severely compromised by non-compliance.

Finally, following Jalan and Ravallion (2003) we also tested for potential remaining selection bias on unobserved characteristics by applying the Sargan-Wu-Hausman test. Specifically, on the sample of those that are part of the common support we ran an ordinary least squares regression of the outcome variable (e.g., the PPVT score) on the propensity score, the residuals from the participation equation, as well as a set of additional control variables that exclude the instrument used (i.e. having an odd or even id code assigned). Since the coefficient of the residuals is not significantly different from zero, there is evidence that remaining selection bias may not be a problem in estimating the impact of the PDAs.

Although we have not been able to identify a convincing explanation for all of the impacts depicted, it is clear that those questions associated with large tables that were split into multiple pages generate a systematic underreporting pattern when the questionnaire was applied using the PDA. For example, in the case of questions related to the socio-economic shocks experienced by any member of the household, the different shocks were divided into eight groups with a leading question at the beginning. Although the structure of the table is identical in both formats, in the case the answer generates a jump or skip the PDA format did not allow the surveyor to see which questions were omitted. It is not surprising that the percentage of households that experienced at least one shock in the last 12 months was an average of 13% higher if the paper-form questionnaire was used. A similar pattern was found when the table related to the child's time use was split in several parts to fit into the PDA screen. Similarly, this was also the case for questions that query whether or not the respondent had needed help for a list of topics including child abuse and family violence.

It was not always the case that the PDA was underperforming. In some cases the format of the paper-based questionnaires may hide a question. This is the case of the question that enquires whether or not the child had missed school due to his work. The question was

placed just below a box and it was inadvertently skipped by several surveyors. This did not happen when the PDA was used.

Regarding the variable child work in the last 12 months, the protocol included some checks to assure that the child was indeed working. The protocol generated the need to go back to a previous question in some cases. Such a move, although possible, was less smooth in the PDA-based questionnaire. This difference may be behind the results obtained here.

Conclusions

This paper uses the Peruvian Young Lives longitudinal sample to assess the extent in which the PDAs may affect the quality of the data collected in socioeconomic surveys. The study finds that in most of the cases there are no significant differences between the results obtained from PDA-based questionnaires and paper-based questionnaires. Given the benefits that the PDA generates in terms of time saving and potential cross checking, it is clear that socioeconomic surveys can take advantage of them without jeopardizing the quality of the data collected.

However, in some cases recording data using PDA-based questionnaires has non-negligible effects so additional care needs to be taken to assure that these problems do not affect data quality. Paper-based questionnaires allow for an overview of large complex tables in one shot, something that is not possible in small form factor like the PDAs. Though, this may be possible when larger devices like tablets become available at reasonable cost in developing countries. We have found that skipping questions embedded in large tables may generate some underreporting, something that was managed better in YL paper-based questionnaires. This may not be an intrinsic advantage of paper-based questionnaires over PDAs but it is something that should be incorporated in an improved version of YL PDA-based questionnaires. More broadly, if PDA is used in a socioeconomic survey that includes

large complex tables, a newer and clever PDA-based questionnaire design needs to be implemented. Eventually, large tables may need to be split and consistency checks should be incorporated.

In addition, our results suggest that when a survey includes a long list of items and this items can be divided in groups it may happen that the surveyor is not able "to see the forest for the trees" when administering a PDA-based questionnaire. This problem does not occur in paper-based questionnaires as surveyor can skim more rapidly and easily throughout a large amount of information. In developing countries this problem can be surmounted once large form factor devices like tablets become available at a reasonable cost.

This paper has also shown some hidden benefits obtained when using PDA-based questionnaires instead of paper-based questionnaires. This is the case of forcing a sequence of questions to avoid inadvertently skipping a particular question. Obviously there is a potential trade-off between imposing a certain structure and allowing for more flexibility that should be taken with caution. Although PDA-based questionnaires may assure a low rate of missing data, one needs to take careful consideration if this is correlated or not with improved data quality. It is clear that forcing surveyors to mark data does not necessarily imply an improvement on data quality.

Beyond the development of questionnaire software, there is a need for building a more complex system that covers all the phases of the data gathering process. Beyond this requirement, there is a need for understanding other aspects of the data collection that are not directly but indirectly associated with the use of the PDAs (Koller, Rennert, & Zettl, 2002). In particular, surveyors need to be trained intensively in the usage of this kind of devices in a way that they can appreciate learning a new technology assuring that it will not affect the rapport with respondents.

Notes

1. For a detailed analysis of the advantages and disadvantages of different methods of data collection see Caeyers and De Weerdt (2010); Hewett, Erulkar and Mensch (2003); Hidalgo-Céspedes, Rosero-Bixby and Antich-Moreno (2007); and, Roberts (2007).

2. See Escobal and Flores (2008) for a detail account of the sampling strategy.

References

Baser, O. (2006). Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health*, 9(6), 377-385.

Becker, T. K., Gries, A., Martin, E., & Bernhard, M. (2010). Experiences with a PDA-based Documentation System in Clinical Research. *Journal of Medical Systems*, 2010, May 19. 1-5.

Bernabe-Ortiz, A., Curioso, W. H., Gonzales, M. A., Evangelista, W., Castagnetto, J. M., Carcamo, C. P., … Holmes K. K. (2008). Handheld computers for self-administered sensitive data collection: A comparative study in Peru. *BMC Medical Informatics and Decision Making*, 8(11), 1-7.

Caeyers, B., Chalmers, N., & De Weerdt, J. (2010). *A comparison of CAPI and PAPI through a Randomized Field Experiment.*

Dale, O., & Hagen, K. B. (2007). Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. *Journal of clinical epidemiology*, 60(1), 8-17.

Escobal, J., & Flores, E. (2008). *An Assessment of the Young Lives Sampling Approach in Peru*. (Technical Note 3). Oxford, UK.: Young Lives, Department of International Development at the University of Oxford.

Fletcher, L. A., Erickson, D. J., Toomey, T. L., & Wagenaar, A. C. (2003). Handheld Computers: A Feasible Alternative to Paper Forms for Field Data Collection. *Evaluation Review*, 27(2), 165-178.

Galliher, J. M., Stewart, T. V., Pathak, P. K., Werner, J. J., Dickinson, L. M., & Hickner, J. M. (2008). Data collection outcomes comparing paper forms with PDA forms in an office-based patient survey. *Annals of family medicine*, 6(2), 154-160.

Hewett, P., Erulkar, A., & Mensch, B. (2003). *The Feasibility of Computer-Assisted Survey Interviewing in Africa: Experience from Two Rural Districts in Kenya.* (Working Paper No. 168). Policy Research Division.

Hidalgo-Céspedes, J., Rosero-Bixby L., & Antich-Montero, D. (2007). Mejora de la calidad y disminución de costos de censos y encuestas utilizando computadores de mano (PDA). Una aplicación en Costa Rica. *Población y Salud en Mesoamérica*, 5(1).

Jalan, J., & Ravallion, M. (2003). Estimating the benefit incidence of an antipoverty program by propensity-score matching. *Journal of Business and Economic Statistics*, 21(1), 19-30.

Jaspan, H. B., Flisher, A. J., Myer, L., Mathews, C., Seebregts, C., Berwick, J. R., … Bekker, L.-G. (2007). Brief report: Methods for collecting sexual behaviour information from South African adolescents--a comparison of paper versus personal digital assistant questionnaires. *Journal of Adolescence*, 30(2), 353-359.

Juniper, E. F., Langlands, J. M., & Juniper, B. A. (2009). Patients may respond differently to paper and electronic versions of the same questionnaires. *Respiratory Medicine*, 103(6), 932-934.

Koller, W., Rennert, F., & Zettl, G. (2002). e-Quest: A metadata-based system for electronic raw data collection. *Statistical Journal of the United Nations Economic Commission for Europe*, 19(3), 187-200.

Lane, S. J., Heddle, N. M., Arnold, E., & Walker, I. (2006). A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Medical Informatics and Decision Making*, 6(23), 1-10.

Lu, Y. C., Xiao, Y., Sears, A., & Jacko, J. A. (2005). A review and a framework of handheld computer adoption in healthcare. *International Journal of Medical Informatics*, 74(5), 409-422.

Rivera, M. L., Donnelly, J., Parry, B. A., Dinizio, A., Johnson, C. L., Kline, J. A., & C. Kabrhel. (2008). Prospective, randomized evaluation of a personal digital assistant-based research tool in the emergency department. *BMC Medical Informatics and Decision Making*, 8(3), 1-7.

Roberts, C. (2007). *Mixing modes of data collection in surveys: A methological review.* (NCRM Methods Review Papers 008). ESRC National Centre for Research Methods.

Shapiro, J. S., Bessette, M. J., Baumlin, K. M., Ragin, D. F., & Richardson, L. D. (2004). Automating research data collection. *Academic Emergency Medicine*, 11(11), 1223-1228.

Stolworthy, Y. (2003). PDAs and Clinical Data Management. *The SoCRA SOURCE Journal*, August, 2003 – 24-26.

VanDenKerkhof, E. G., Goldstein, D. H., Blaine, W. C., & Rimmer, M. J. (2005). A Comparison of Paper with Electronic Patient-Completed Questionnaires in a Preoperative Clinic. *Anesthesia & Analgesia*, 101(4), 1075-1080.

Tables

Table 1. Average effect of the treatment on the treated (ATT) under different matching estimators. Younger cohort – Round 3.

Variable	Mean comparison test	Matching Technique		
		Nearest Neighbor	Stratification	Kernel
Household Questionnaire				
Mother didn't migrate since the last interview	**		*	
Mother thought about moving away from community		*		
The Child is not attending to school	*		*	
Child has any difficulties in getting to school		**		
Hours per day child spent on household chores			*	*
Hours per day child spent on unpaid work		**	*	*
Hours per day child spent on paid work		*		
Hours per day child spent playing	***	*	***	***
Help received from caregiver's parents in case of problems			**	***
Taken part in a protest march	*	*	**	**
Juntos Program is good/ very good		*		
Household experienced at least one shock	***	***	***	***
Asset value at median prices	*			*
Days per week were the child is physically active		*	**	***
Children were affected for food problems	*			
Child Questionnaire				
Child is doing better compared to other children in class		**		
Child has missed school due to child work	***	***	***	***
Child work in the last 12 months		*	**	*
Hours per week spent on the internet	*		**	*
Child feel responsible for his pet		*		
Degree of Risk Aversion		*	**	**
Number of estimated outcomes	*60*	*60*	*60*	*60*
Number of significant outcomes	*9*	*14*	*13*	*12*
Percentage of significant outcomes	*15%*	*23%*	*22%*	*20%*

Note: Differences are significant at ***1%, **5%, and *10%.

Table 2. Average effect of the treatment on the treated (ATT) under different matching estimators. Older cohort – Round 3.

Variable	Mean comparison test	Matching Technique		
		Nearest Neighbor	Stratification	Kernel
Household Questionnaire				
Double residence			**	
Having a dictionary at		**		
Hours per day child spent caring for family members	**			
Hours per day child spent on household chores	**			
Hours per day child spent on paid work	*		*	
Hours per day child spent at school		*	*	*
Hours per day child spent studying		**		
Hours per day child spent playing	*			
Help received from caregiver's parents if problems		**		
Taken part in a protest march	*	*	*	**
Use of internet by household members	**	**	**	**
Access to JUNTOS	*		**	**
Asking for help (child abuse or family violence)	**	**	**	**
Household experienced at least one shock	**	*	*	**
Use MINSA's services when child is ill?		*		
Child has DNI	*		*	
Child is registered in SIS		***	*	*
No food problems	*			
Household members smoke cigarettes	**	**		
Child Questionnaire				
Hours per day child spent on unpaid work	*			*
Hours per day child spent at school				*
Child work in the last 12 months	**	**	***	***
Child can leave his job if he don't get paid on time	**	**	***	***
Household is rich/very rich	*			
Child currently enrolled in school	*			
Hours per week spent on the internet	*			
Number of estimated outcomes	*81*	*81*	*81*	*81*
Number of significant outcomes	*18*	*13*	*12*	*11*
Percentage of significant outcomes	*22%*	*16%*	*15%*	*14%*

Note: Differences are significant at ***1%, **5%, and *10%.

Table 3. Average treatment effect (ATE) of selected outcomes. Regression model without interactions. Younger cohort – Round 3.

Variable	ATE	
Household Questionnaire		
Hours per day child spent playing	-1.13	***
Household experienced at least one shock	13.87	***
Asset value at median prices	143.60	*
Children were affected for food problems	-3.30	*
Child Questionnaire		
Child has missed school due to child work	4.90	***
Number of estimated outcomes	60	
Number of significant outcomes	5	
Percentage of significant outcomes	8%	

Note: Differences are significant at ***1%, **5%, and *10%.

Table 4. Average treatment effect (ATE) of selected outcomes. Regression model without interactions. Older cohort – Round 3.

Variable	ATE	
Household Questionnaire		
Hours per day child spent on household chores	-0.93	**
Hours per day child spent on unpaid work	1.53	*
Hours per day child spent at school	0.93	**
Hours per day child spent studying	0.86	*
Hours per day child spent playing	-0.20	*
Taken part in a protest march	-32.25	*
Use of internet by household members	6.66	***
Access to JUNTOS	19.63	*
Asking for help (child abuse or family violence)	-16.25	**
Household experienced at least one shock	63.38	***
The child has DNI	-13.52	*
The child is registered in SIS	58.52	**
Household members smoke cigarettes	2.96	*
Child Questionnaire		
Hours per day child spent on unpaid work	-0.07	*
Child work in the last 12 months	-0.07	**
Child can leave his job if he don't get paid on time	-0.08	**
Number of estimated outcomes	82	
Number of significant outcomes	16	
Percentage of significant outcomes	20%	

Note: Differences are significant at ***1%, **5%, and *10%.

Table 5. Average treatment effect (ATE) of selected outcomes. Interaction model. Younger cohort – Round 3.

	Transmission Channels									
	Small sample estimation bias		Field worker		PDA		PDA ATE			
	Chi-squared reported									
Household Questionnaire										
Hours per day child spent playing	9.33		180.40	***	80.85	***	-0.35	***		
Household experienced at least one shock	15.44		176.70	***	73.53	***	-45.25	***		
Asset value at median prices (soles)	5.26		28.50		24.21		-641.10			
Were the children in the household also affected for food problems?	13.38		66.17	***	20.86		-3.41			
Child Questionnaire										
Has missed school due to child work	42.92	***	19.15		109.70	***	4.48	***		

Note: Differences are significant at ***1%, **5%, and *10%.

Table 6. Average treatment effect (ATE) of selected outcomes. Interaction model. Older cohort – Round 3.

	TRANSMISSION CHANNELS							
	Small sample estimation bias		Field worker		PDA		PDA ATE	
	Chi-squared reported							
Household Questionnaire								
Hours per day child spent on household chores	14.22		42.64	**	41.15	**	-0.43	**
Hours per day child spent on unpaid work	17.26		38.28	*	26.47		1.61	
Hours per day child spent at school	10.56		37.63	*	28.13		1.60	
Hours per day child spent studying	16.18		36.22	*	39.82	*	-0.09	*
Hours per day child spent playing	23.61	**	64.76	***	40.76	**	-0.22	**
Taken part in a protest march	24.35	**	25.75		36.64		-67.28	
Use of internet by household members	18.68		34.72		40.57	**	2.11	**
Access to JUNTOS	12.71		66.15	***	34.72		-18.44	
Asking for help (child abuse or family violence)	14.13		43.86	**	41.56	**	-62.78	**
Household experienced at least one shock	13.85		64.76	***	33.61	**	26.24	
The child has National Identity Card	25.88	**	54.77	***	53.48	***	-71.62	***
Child is registered in Health System (SIS)	27.01	**	36.48	*	55.89	***	34.03	***
Household members smoke cigarettes	18.23		24.59		28.61		10.75	
Child Questionnaire								
Hours per day child spent on unpaid work	24.31	**	43.06	**	37.16	*	0.03	*
Child work in the last 12 months	15.52		114.20	***	53.74	***	-11.46	***
Child can leave his job if he don't get paid on time	15.27		36.48	*	34.10		41.35	

Note: Differences are significant at ***1%, **5%, and *10%.