Individual measure of reliability: Evidence from Malawi's MDICP Project

Ruben Castro[1]

Abstract

Empirical work in social sciences relies extensively on survey data, raising considerable interest on the quality of that kind of data. According to literature, not random noise is found in survey data, resulting from factors as interviewer effects, question wording, social desirability and others. Rather than question-individual level, in this paper I propose a measure of reliability at individual level, and show how this measure is useful for research on the process of answering a survey. I compute the reliability index from the most simple questions contained in the MDICP survey in Malawi, and interpret it as the cognitive resources invested in answering the survey; then I show how unreliable individuals provides statistically different answers than other people in virtually any kind of questions, even after controlling for stratification variables. As an application, I propose a simple model to show how couples' disagreement about household's wealth is in part explained by the reliability index, where both men and women unreliability is associated with higher reports.

Introduction

Empirical work in social sciences relies extensively on survey data. Several studies, however, indicate the presence of random and non random errors in the data (for early evidence, see Freedman and Sun 1974, Koenig et al. 1974, Bound 1989) resulting from a wide set of circumstances such as lying, memory limitations, economic interests, cognitive issues (Harding 1991), interviewer effects, and of course issues pertaining to question design

---

[1] Demography PhD student, University of Pennsylvania. This article receive considerable help from Susan Watkins.

(Biemer 2003). The literature so far has concentrated on detecting and exploring the genesis of response error, leading to findings about, for example, questions interpretation (Smith and Morgan 1994 expose this issue in a mother - daughter framework), interviewer effects (Bignami 2003), cognitive problems, social desirability influences, and subjective questions; Bertrand and Mullainathan 2001 present a nice summary of subjective questions, emphasizing that *perhaps the most devastating problem with subjective questions, however, is the possibility that attitudes may not exist in a coherent form*

In this paper I propose a measure of reliability at individual level, which I compute from the most simple questions contained in a survey, and interpret it as the cognitive resources invested in answering the survey. The survey used in this article is the *Malawi Diffusion and Ideational Change Project* (MDICP); Crude inconsistency in MDICP survey ranges around a crude 10% of answers being inconsistent between wave 1 and 2, and this percentage goes up to 20% and 30% when referring to family planning and sexual behavior questions (evidence from Bignami 2003). Using Kenyan data, Watkins 2001 founds 30% of responses were inconsistent between husband and wife, while Bignami 2004 reports a MDICP crude couple inconsistencies as high as 30 to 40% for questions related to "family planning", "fertility planning" and "discussions about number of children", compared to 10% inconsistency in "assets" questions.

However, consistency analysis has so far concentrated on respondent-and-question level, and also at question level, rather than analyzing an individual level measure of consistency; this measure may help researchers working on the cognitive proves of answering and may also be used to improve estimations. In regression analysis, for example, the most common

way to acknowledge consistency issues in data analysis is through measurement error theory (Bound, John, Charles Brown, and Nancy Mathiowetz 2001 presents an impressive in this topic), where proposed methodologies to attenuate the impact of measurement error require adding a considerable level of complexity, as well as extra assumptions, in already complex models. Thus in this framework, researcher might be discouraged of considering some reporting error on their estimates.

But several surveys are collected in two or more waves, making possible to compute credible individual measures of consistency, suitable to be used elsewhere. And using an estimated consistency measure, built at individual level, may provide an alternative way to include data quality issues in regular data analysis, including the study of survey design. As an example, if consistent individuals exhibit, conditional on their observed characteristics, smaller random error in their answers, but not bias, an heteroskedasticity function based in the reliability index might improve the efficiency of the estimates. Also, bigger weights could be placed on consistent individuals, but this procedure would probably bias the estimates toward them. Also, the reliability index could somehow be included in regression analysis, attempting to correct for reporting error; Watkins et al. 2001 claims that their *analyses demonstrate that discrepancies between the reports of husbands and wives sometimes have a systematic gender component that affects staple questions of many surveys, such as those concerning household possessions*. How much of the systematic bias can be attributed to our measure of reliability?.

This research proposes an individual consistency measure, and then focuses on its descriptive analysis using empirical data from Malawi's MDICP. As an application, I propose a simple model for wife and husband discrepancies on economic conditions. Smith et all

1994 shown how both question and individual characteristics influence couple disagreement; in this article I model the deviation from the true, and I show how a individual characteristic in particular, the reliability index, is correlated with this deviation. Interestingly, each spouse's reliability has a positive correlation with the deviation, thus leading to the conclusion that both men and women unreliability is associated with higher reports.

Framework

In the answers to any particular question some error may arise either from individual, question, or individual-question specific effects; this research focuses on individual effects. The kind of individual effect studied here might be used to study several issues related to the cognitive process of answering. Naturally, this analysis is far from straightforward and, though some ideas are given in the text, I leave it for future research.

The proposed strategy relies on the individual response theory, IRT, which in short words assumes a latent variable at individual level, different properties for every question considered and a random component in the answers. In the psychometric literature, IRT is usually applied to recover the unobserved *ability* of individual, using data on several questions; IRT assumes that questions involved in the analysis have an inherent level of difficulty while individuals have an inherent level of ability (Skrondal 2004 provides a general framework). Interestingly, IRT theory have been also extended to health analysis (cook et al. 2007, Hays et al. 2007).

In order to produce a standard interpretation index, similar methodology will be used in this research to recover a close parallel of ability. The proposed "latent variable" is the *cognitive*

resources invested in answering the survey, whereas it comes on purpose (individual decide not to bother in giving a good answer) or it is just the outcome of the individual cognitive skills. So the consistency measure will reflect the *cognitive resources invested in answering the survey* (in all the following text consistency refers to this specific kind of consistency). As the reader might note, a key assumption here is that questions included in the analysis are reasonably free of any kind of influence, for example given by social desirability.

Assuming we have an individual consistency measure, the joint distribution of this index and the observed characteristic of individuals is a natural second step. Do low consistency individuals report themselves consistently different than the average individual? Preliminary data seems to support that: it shows that low consistency individuals report less friends dying because of HIV, and men and women over and under reporting assets, respectively.

Why the association between our reliability index and any kind of question included in a survey? Although more consistent individuals invest more cognitive resources in answering the survey, this is not the only phenomena going on, and is beyond the scope of this article to explore in detail the underlying phenomena behind the joint distribution of the index and other data from the survey. For example, let's consider the reports on extra marital affairs. Reliable individuals, in the sense of reliability proposed in this articled, just invest more cognitive resources in coming up with what they consider is the right answer; this might imply a truthful answer, or a better lie if they want to lie.

The data used in the article is MDICP, which had been collected in some rural districts on the south, central and north area of Malawi, in the years of 1998, 2001, 2004, 2006 and 2008,

though the last wave is not included in this article. MDICP had been already subject to different to response error analysis, as for example Bignami 2003a, using a sub sample of individuals which were interview twice and Bignami 2003b, analyzing interviewer effects, question reliability and sample attrition.

Last section presents an application of the index, proposing a simple model for couple disagreement on household reports. The analysis appears to be helpful in disentangling male and female contributions to the overall couple-inconsistency.

Section 1: The reliability scoring

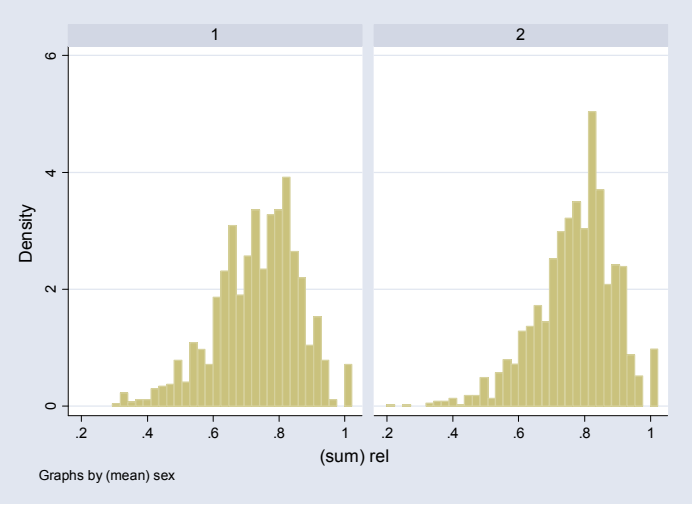
Each time a question is asked in two consecutive waves, it may be possible to check consistency if the answer on wave 2 should be correlated with the answer on wave one. To compute the reliability score, we selected in particular "plain" questions, meaning questions where we expect a low level of pollution from question-specific bias. Among our plain questions we include "year born", "ever been to school", "year at first marriage" and "wherever that marriage ended", "ever been outside the country", "number of children", "number of dead children", and others. The reliability score is computed just as the ratio between consistent answers over all answers measured, so for called *check points*. The original idea was to use the IRT framework to recover the latent reliability, but just the simple score works well for the purposes of this paper; the simple score might be viewed from the perspective of assuming that every question have the same properties and that the size of the "mistake" have a linear relation with the latent variable.

So far we are using check between waves 2001/1998, 2004/2001 and 2006/2004. As can be seen in the next table, the reliability 2001-2004 is the one with less check points, due to lack of comparability between questionnaires. Also, the reliability is somewhat lower in this case; this may be due to several factors, maybe related to the kind of questions available, differences in the survey management or others:

	Mean number of check points		Mean reliability	
Waves	Men	women	Men	Women
1998-2001	13	11	81%	81%
2001-2004	9	6	70%	72%
2004-2006	13	12	72%	76%

Also, we can appreciate a slightly bigger level of female consistency, an outcome that show up repeatedly in the consistency statistics. Adding all the check points together, next graph shows the histogram of the reliability score by gender:

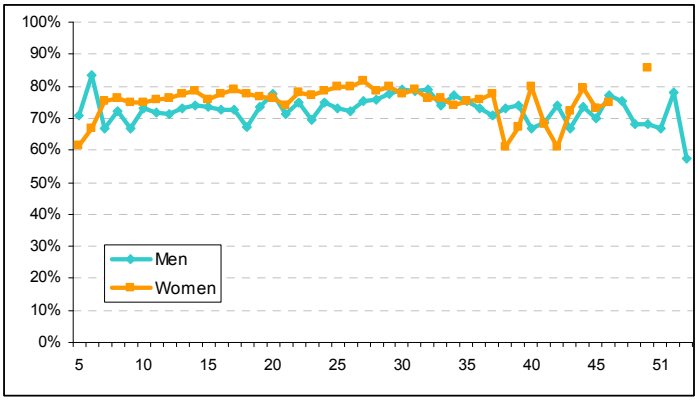
Histogram of reliability, by gender



An important issue is the different set of check points available for each individual: some questions are only observed for a sub set of the individuals, thus selection issues may influence results. For example, number of dead children can not decrease between waves, so

women in high mortality household are more likely to pass this check point, whether if comes from true reliability or from the high mortality, which mask unreliable reports. It is in the hope of this article that once the different check points are added together, different selecting issues start to cancel themselves out. Related with this last point, if the questions included in the analysis are truly reflecting the reliability concept proposed in this article, the number of check points available for each individual should not be correlated with the reliability index; although some form of correlation may be visible from the previous table, next figure shows that is not the case:

Mean reliability, by number of check points and gender



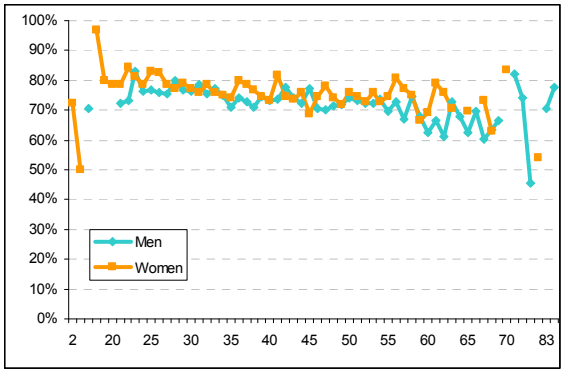
To compute the previous graph, and in general to carry on the whole study, we are selecting only individuals with at least five check points, hoping that different selection issues tend to cancel themselves if we have more than a minimum level of check points, chosen qualitatively to be five.

Next table shows the aggregate final statistics:

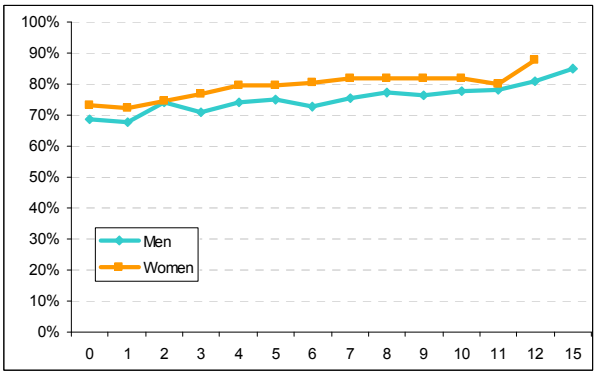
	Men	Women
Mean number of checks	22.8	20.4
Mean reliability	74%	77%

If reliability is capturing cognitive resources invested in answering the survey, there are some reasons to expect a negative relation between age and reliability: memory and mental skills for the involuntary side, and feeling uninterested in the survey from the voluntary side. However, next figure shows a somewhat slowly decreasing of reliability with age. In the case of education, we also intuitively expect it to be positively association with the reliability index, as is also shown in next figure:

Mean reliability by age and gender[2]



Mean reliability by education and gender



*figures computed using age and education reported in 2001. Sample size is extremely low over 12 years of education.

As a word of caution for interpretations, it is necessary to emphasize the overall low level of education of this population, especially for women, as reported by the individuals:

	Men			Women		
Years of education	1998	2001	2006	1998	2001	2006
0	174	163	170	466	484	454
1-3	170	206	204	260	329	292
4-7	212	204	214	329	342	350
8	161	197	170	160	174	145
8+	113	114	135	61	83	75
TOTAL	*830*	*884*	*893*	*1,276*	*1,412*	*1,316*

[2] Distribution of individuals by age in appendix A

Section 2: Joint distribution of reliability and selected answers

This section shows how the answers in the survey show a consistent correlation with the reliability index. Several questions are included in more than one wave (1998, 2001, 2004 and 2006), among which we selected the questions included in the index. In the following, the answers for selected questions are tabulated according to the reliability quintile of the respondent.

Very important note: in this section each selected question is also regressed on reliability, age and education, to confirm that conclusions appreciated in the tables are not just a mirror of stratification, or some artifact of the underlying structure of age and education. If still the reliability index is reflecting some form of stratification, rather than cognitive resources invested in answering the question, is discussed in the concluding section.

Condoms

As it is the case with question about HIV, condoms is a sensible issue in Malawi, so the underlying relation of condom reports and reliability may be very complex and perhaps not visible in the tabulations. Even in the case of clear correlation, interpretation is not straightforward. Nevertheless, reliability and condom acceptance show an interesting pattern in the data:

It is acceptable to use a condom with the spouse?

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1: Lowest quintile	15%	31%	29%	28%	19%	28%	36%	40%
2	12%	24%	32%	30%	18%	29%	44%	40%
3	7%	18%	24%	26%	13%	28%	38%	40%
4	10%	20%	26%	28%	12%	29%	40%	37%
5: highest quintile	5%	23%	20%	24%	13%	32%	41%	35%

In the case of men, it seems that more reliable individuals report less acceptance of condom use with spouse, in all waves. In the case of women the evidence is less clear, except for waves 1998 and 2004.

In the case of women, though, we get a much clear relation when the question refers to *actual facts* about condom use. In particular, reliable women report more experience in condom use (the highest quintile of reliability report around twice as much experience with condoms):

	Have you ever use a condom? (women)		Are you now using condom? (women)		
Reliability	1998	2001	1998	2001	2004
1	1.1%	5.7%	1.9%	2.2%	8.3%
2	3.5%	5.0%	3.1%	4.8%	4.2%
3	3.6%	9.8%	1.3%	5.8%	13.9%
4	1.3%	9.2%	1.1%	5.7%	6.3%
5:	8.6%	10.5%	6.9%	9.3%	14.1%

Children

Reliable individuals, both men and women, in every wave, reported smaller numbers of children ever had. Furthermore, the gradient is very strong and the difference between the lowest and highest quintile is around 1 child.

How many kids do you ever have?

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	5.1	5.6	5.7	5.8	4.6	5.1	5.3	5.1
2	4.8	5.3	5.4	5.7	4.5	5.0	5.0	4.8
3	4.7	5.4	5.6	5.9	4.3	4.9	5.1	4.7
4	4.1	5.1	5.2	5.6	4.0	4.6	4.7	4.1
5	3.7	4.7	4.6	5.1	3.7	4.5	4.7	3.7

In the case of living children (next table), we appreciate a different picture for women, though not for men: reliability quintile doesn't show a correlation with women report on

living number of children. If the true number of living children is actually not correlated with the reliability index, perhaps close ties between mother and living children may easier to come up with the right answer. This argument is of specially interest in arguing that the reliability index is not just the mirror of some form of stratification, at least in the relation to family size.

How many living kids do you have?

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	4.4	4.5	4.9	5.1	3.5	3.7	4.0	4.3
2	4.3	4.4	4.8	4.9	3.7	3.9	3.9	4.2
3	4.0	4.4	4.9	5.1	3.6	4.0	4.1	4.4
4	3.8	4.2	4.8	4.9	3.3	3.7	4.0	4.3
5	3.5	4.0	4.0	4.6	3.2	3.7	4.0	4.4

HIV

As it was said for condom reports, interpretation of the following table is not straightforward. Again, the point of this section is to show how the reliability measure used in this document is also involved in the process of answering.

How many people know to you have died from HIV, overall?

	Men		Women	
Reliability	2001	2006	2001	2006
1	7.0	7.5	5.1	6.3
2	8.7	9.9	6.6	7.9
3	7.7	9.5	7.1	8.1
4	9.5	10.0	8.4	10.2
5	10.8	11.3	9.7	10.6

So, reports on this question can grow between 50% and 90% when reliable individuals are answering. To attenuate the recall bias, we compute this table with a top coding on numbers bigger than ten, with similar results. We also tabulate a table for the proportion of individuals

reporting at least one people died from HIV, and we found results are just in line with the previous table:

Proportion reporting at least one: How many people know to you have died from HIV?

	Men		Women	
Reliability	2001	2006	2001	2006
1	96%	95%	91%	93%
2	98%	98%	96%	94%
3	98%	96%	97%	94%
4	99%	97%	98%	97%
5	100%	98%	97%	98%

Another sensible question included in the surveys is related to relatives dying from HIV. This question should also be less affected by recall bias. Again, reliable individuals tend to report higher numbers, with an impressive 100% difference between quintile 1 and 5.

How many of your relatives people said had died from HIV?

	Men		Women	
Reliability	2001	2006	2001	2006
1	0.8	1.2	0.8	1.0
2	1.2	1.4	1.0	1.3
3	1.1	1.4	1.2	1.5
4	1.2	1.6	1.3	1.8
5	1.3	2.0	1.4	2.1

Sexual behavior

Sexual behavior is asked in all waves, though the kind of questions had been changing. Three facts about personal sexual behavior where ask in the 2001 wave, most of them showing a somewhat weak correlation with the reliability score, clearer in the case of women (see next table). Perhaps unexpectedly, reliable women report higher age at first sex, higher marriage rate with the first partner and also, conditional on breaking up, shorter relationship duration.

Though the latter result may be more understandable, the first ones come at odds with conventional wisdom. A possible explanation is the different age and education structure behind the reliability level, but a regression analysis shows a strong significance for the reliability coefficient, in the case of women, for both age at first sex, and married rate.

	Age at first sex		Married the first sexual partner		Years with the first sexual partner*	
Reliability	Men	Women	Men	Women	Men	Women
1	17.8	16.2	35.1%	61.6%	1.79	1.84
2	17.9	16.2	30.1%	61.9%	1.69	1.57
3	18.1	16.6	38.4%	67.3%	1.66	1.51
4	18.8	16.8	29.9%	64.1%	1.83	1.45
5	18.0	16.9	24.6%	73.2%	1.91	1.49

*conditional on not being still together; this question has a small sample size (50)

Another sensitive question is total number of sex partners, a question included in the last two waves. On average, men reported a total of 3.7 whereas women reported 1.9. Regardless of the accuracy of the reports, reliability was strongly correlated with answers, especially for women.

How many sex partners had you ever had?

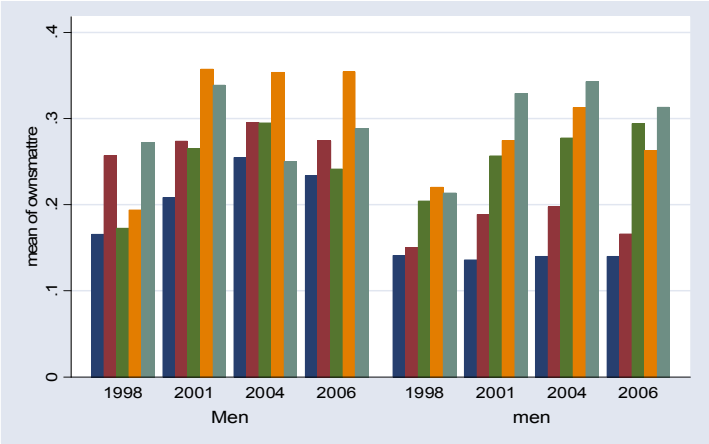
	Men		Women	
Reliability	2004	2006	2004	2006
1	5.2	5.5	2.5	2.3
2	4.6	4.5	2.1	2.1
3	3.6	3.9	1.8	2.0
4	5.1	4.2	1.7	1.9
5	3.6	4.8	1.4	1.7

Again, results might seem at odds with conventional wisdom: both reliable men and women recognize fewer sex partners. However, in explaining these findings it is necessary to remember that the conceptual basis of the index is related to cognitive resources invested in answering the survey, whether the individual wants to build a good report or a good lye.

Economic conditions

The following are the main tabulations for economics question. Basically, in almost every economic question, women report higher economic conditions the higher the reliability measure. In the case of men, the situation is similar, though somewhat less clear; for example:

Does the HH own a mattress?



*x's: reliability quintile; y's: mean answer

Personal income

Among individuals reporting having received some wage, and in particular reporting it in an annual rate (the most common answer), the following table shows a continuous gradient for both men and women.

Wage in current work, in an annual basis (in 00 of Kwacha):

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	75	122		124	44	57		65
2	73	146		111	39	68		67
3	65	145		114	46	92		109
4	88	172		121	42	89		103
5	90	156		152	62	123		82

*100 = US$650 aprox.

Wage in last week/month (in 00 of Kwacha):

	Last week				Last month	
	Men		Women		Men	Women
Reliability	1998	2001	1998	2001	2001	2001
1	11	33	3	16	22	12
2	8	33	3	16	28	15
3	14	34	7	19	27	15
4	23	45	6	20	37	15
5	19	35	6	22	27	21

Though not shown here, percentage reporting been in a paid work has also a gradient for women; in the case of men, basically 100% of men report "yes" at every reliability level. About percentage reporting wage in the last week/month, there is also positive correlation with reliability, for both genders.

MDICP also contains reports on total personal and household spending in clothes, medicine and others:

How much is your personal spending?

	% Reporting positive spending				Reported $ among			
	Men		Women		Men		Women	
Reliability	2004	2006	2004	2006	2004	2006	2004	2006
1	73%	50%	69%	55%	8.8	11.6	5.4	8.6
2	81%	63%	77%	55%	8.1	10.1	5.7	8.4
3	74%	55%	74%	61%	8.1	10.1	5.8	8.9
4	80%	62%	78%	66%	7.6	11.7	6.2	7.9
5	66%	58%	79%	60%	9.0	12.4	6.5	8.8

Household assets

Once again, the reliability index shows a correlation with the answers; in this case, the HH assets. To summarize results about HH goods, following table shows aggregate reports on number of goods, including: mattress, bicycle, radio, pit latrine and paraffin lamp, because this goods had been included in every wave:

Number of HH goods reported

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	1.9	2.3	2.2	1.5	1.8	1.8	1.8	1.8
2	2.3	2.6	2.6	1.6	2.0	2.1	2.1	2.1
3	2.3	2.5	2.4	1.9	2.3	2.3	2.3	2.3
4	2.6	2.7	2.8	1.9	2.3	2.6	2.5	2.5
5	2.5	2.3	2.4	1.9	2.6	2.7	2.4	2.4

About HH farm animals, next table summarize the total number of animals among cattle, goats, pigs and poultry. Surveys ask separately for each kind of animal, and just for the sake of simplicity we collapse them all in one variable. Nevertheless, conclusions are the same for each kind of animal.

Proportion reporting any kind of HH farm animals

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	55%	63%	65%	60%	54%	71%	55%	59%
2	65%	70%	74%	74%	62%	71%	68%	65%
3	65%	72%	72%	67%	63%	76%	65%	65%
4	68%	77%	75%	74%	67%	76%	71%	71%
5	68%	69%	60%	64%	67%	82%	73%	67%

If farm animal were reported, how many the HH owns?

	Men				Women			
Reliability	1998	2001	2004	2006	1998	2001	2004	2006
1	11.1	12.2	11.5	11.1	10.1	9.3	9.9	8.4
2	12.1	13.6	10.5	12.7	10.3	10.1	12.1	10.2
3	11.0	13.2	13.8	12.4	10.3	11.2	11.2	11.1
4	14.2	13.5	12.1	12.2	11.3	12.8	12.1	11.2
5	13.7	13.4	14.6	13.0	12.4	13.2	13.1	11.0

The last variable is HH purchases, which is shown in the next section, where if is used to analyze couples disagreement.

Section 3: reliability index applied to couples inconsistency

The remaining of the paper focuses on couples reports, in particular in analyzing couples disagreement about HH economic situation. This area where chosen because the "social desirability" ingredient may be easier to interpret and because the regression analysis is less challenging in the case of continuous variables.

In general, couples disagreement about HH's economic conditions follow a pattern where men reports are higher, also found in Watkins et al 2001. This had lead to the conclusion that men like to look "strong", overemphasizing the economic situation of the household, whereas women try to look "needy", under reporting the economic conditions, in the hope, presumably, of getting some support. With the reliability index proposed in this article, cognitive resources invested in answering the survey, we can not distinguish the accuracy of this hypothesis because there is probably more elements involved in the disagreement, but we can determine whether the reliability plays also a role in the disagreement.

Nevertheless, if the index plays similar than that of Watkins proposal, men's reliability should be negatively correlated with their reports, everything else equal, whereas the opposite should be the case for women. Previous tabulations contradict this hypothesis, as we see that both sexes show positive correlation reliability/assets, but this evidence is, unfortunate, insufficient: there may be some underlying differences correlated with reliability. For example, more reliable individuals may be richer, so men and women will show positive correlation between reliability and economic reports, regardless of the "strong" and "needy" attitudes of the spouses.

However, the positive correlation between reliability and economic reports remains significative after controlling by age and education, so a natural conclusion is that unreliable individual pretend to be poorer than what they actually are. If education is not a strong proxie for the true underlying resources, though, this conclusion may not hold. Following methodology attempts to get rid of the unobserved "true" answer in order to analysis the relation between misreports and reliability.

Methodology

The only continuous variables related to HH economic conditions is, in wave 2004 and 2006, spending of the household. For simplicity, just summation of reported spending is going to be used in this section. Other options, for future research, are some discrete variables, as ownership of different kind of animals, agricultural production and household goods.

So the variable used in the analysis is *how much is the total spending of the household? (Wave 2004)*, which is asked separately by cloths, medical, fertilizer and other items. This variable shows no clear correlation for men:

How much is the HH spending?:

	% Reporting positive spending				Reported $, conditional on >0			
	Men		Women		Men		Women	
Reliability	2004	2006	2004	2006	2004	2006	2004	2006
1	71%	67%	64%	68%	19	29	11	18
2	77%	80%	73%	74%	20	26	11	17
3	78%	71%	72%	74%	18	26	14	21
4	79%	78%	76%	78%	23	29	15	21
5	66%	74%	79%	73%	20	27	16	22

For the purpose of couples analysis, only non-polygamous men are going to be considered, which comprise around 90% of the total. The average couple reports show a clear correlation with reliability, which is the outcome of a) a true gradient in wealth and some form of stratification, of which reliability acts as a proxy and b) a true causation from reliability to the given answer.

Average couple-report by reliability interval of husband and wife

		Wife's reliability interval				
		1	2	3	4	5
Husband's reliability interval	1	1,342	1,555	1,409	2,311	2,268
	2	1,256	2,087	2,604	1,558	4,529
	3	1,189	2,063	2,092	1,177	1,971
	4	2,155	3,782	1,944	4,290	2,350
	5	2,571	1,493	4,004	3,126	3,261

Now, about the difference between spouses, next table show the average difference over the average answer:

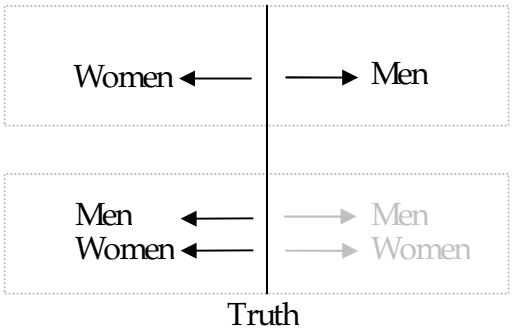
Average disagreement over average answer

		Wife's reliability interval				
		1	2	3	4	5
Husband's reliability interval	1	0.52	0.42	0.06	0.45	0.56
	2	0.48	0.27	0.36	0.63	0.30
	3	0.75	0.61	0.20	0.13	0.62
	4	0.24	0.05	0.29	0.33	0.56
	5	0.14	0.32	0.20	0.22	0.21

As can be seen from a column by column inspection, there isn't a clear pattern in the case of women, and there is some pattern for men: bigger reliability look somewhat related with smaller disagreement/average answer.

The model

The hypothesis we want to test here is that reliability, as conceptualized in this article, is influential in the couple disagreement, and, whether the association between reliability and misreport have the same sign in men and women:



Hip 1: Men over report/ Women underreport

Hip 2: Both misreport in the same direction

To estimate whether men and women are or not misreporting, we will define:

$$R^{spouse} = truth + mistake^{spouse} + random$$

where spouse may refer to the husband or the wife. This equation is very unfortunate because we don't observe any of the x' components: the truth answer, the lying or mistake component, and the random error. However, we will assume the "mistake" can be decomposed in a reliability-related and a not reliability component:

$$R^{spouse} = truth + \beta_0^{spouse} + \beta_1^{spouse} * rel^{spouse} + random$$

So we already assume a linear relation between our reliability measure and the reported answer. Finally, we will subtract the wife answer from the husband answer to get rid of the true answer and this is going to be the final equation:

$$R^{men} - R^{wom} = (\beta_0^{men} - \beta_0^{wom}) + \beta_1^{men} rel^{men} - \beta_1^{wom} rel^{wom} + e =$$

$$\boxed{\Delta^{couple} = \beta_0^{couple} + \beta_1^{men} rel^{men} - \beta_1^{wom} rel^{wom} + e}$$

This regression doesn't allow us to estimate separately the non-reliability related lies: we just get the beta zero, which correspond to men minus women; but it does allow us to estimate beta ones. The main convenience of this specification is the non-dependence on the truth level of spending, which is not observed. Results are show under the column MODEL 1:

Dependent variable: household disagreement (actual value, not absolute value!)

	Model 1		Model 2		Model 3	
	Parameter	P> |t|	Parameter	P> |t|	Parameter	P> |t|
Unreliability of men*	1301	(0.111)	1.19	(0.011)	1504	(0.044)
Unreliability of women*	-1635	(0.042)	1.92	(0.000)	-1522	(0.047)
Educ high, men	17	(0.945)	-1.73	(0.218)	10	(0.028)
Educ med, men	-404	(0.149)	-1.34	(0.617)	-143	(0.573)
Educ high, wom	231	(0.531)	-1.11	(0.520)	-348	(0.273)
Educ med, wom	1299	(0.031)	0.21	(0.273)	-22	(0.95)
Average answer					837	(0.122)
Constant	702	(0.015)	0.53	(0.004)	258	(0.34)
N	548		542		541	
R^2	0.03		0.03		0.15	

*Unreliability=1-reliability

From model 1, we can conclude that the influence of men unreliability is almost significant, whereas women' one it strongly significant. The positive and negative sign of the parameter for men and women respectively means that the disagreement tend to grow with reliability, for both sexes, given that women enter the equation with a minus sign. But before interpreting further the results, I'll introduce model 2 and 3.

First issue is that the level of disagreement might be better understand in relative terms, i.e., divided by the true level. This variable is not observed so as a proxi I use the ratio disagreement over couple's report as independent variable, whose results are contained in Model 2. Also, as said previously, we can not rule out a correlation between reliability and some underlying stratification, even if we control for education and even when I got rid of the true answer which is likely to be correlated with stratification. The reason is that the level

of the misreport might be itself correlated with some stratification variable, for example, richer couples with the same relative misreporting than poor couples will have higher absolute disagreement. To deal with this, I include in the regression another independent variable: the average spending reported by the couple, to try to correct for the underlying stratification of the individual; results are shown in Model 3. I also run separate regression for different quintiles of reliability, but these results are not shown here.

From previous table, we appreciate how coefficients maintain their values and gain significance as we include the "average answer" to either use relative disagreement as independent variable or to control for the underlying level of stratification. Results show a positive coefficient for husbands and a negative one for wives. In other words, regression coefficients show that both spouses misreport in the same direction and, perhaps strikingly, given the configuration of parameters, this regression implies that both spouses tend to over report as they get less unreliable.

Conclusion

Cognitive resources invested in answering a survey matter. A simple score of consistency between waves, computed using only "plain" question, show how unreliable individuals provides statistically different answers than other people in virtually any kind of questions, even after controlling for stratification variables.

Why some people invest more resources than others? Why the correlation between any given question and the reliability index show the patterns we have seen in the document? .

this are questions that might be complemented with the evidence shown in this article. In the case of

As an application, I propose and estimate a simple model to show how couples' disagreement about household's wealth is in part explained by the reliability index. The regression gives stable values of the reliability coefficients, showing that both men and women unreliability is associated with higher reports. This result is not directly comparable with Watkins et al 2001, who shows that men on average reports greater economic condition than women, but if we assume that reliable individuals report more accurately it is possible to conclude that both spouses over report, but men do so more aggressively.

References

[1]. J Bound, A. Krueger. "The Extent of Measurement Error In Longitudinal Earnings Data: Do Two Wrongs Make A Right?.NBER Working Paper, 1989.

[2]. FJ Fowler, TW Mangione. "Standardized Survey Interviewing: Minimizing Interviewer-Related Error", 1990.

[3]. Paul P. Biemer, Lars E. Lyberg. "Introduction to Survey Quality. Series: Wiley Series in Survey Methodology". Published Online: 27 Mar 2003.

[4]. Bignami-Van Assche, Simona. 2003. "Are we measuring what we want to measure? Individual consistency in survey response in rural Malawi," Demographic Research S1(3): 77-108.

[5]. Bignami-Van Assche, Simona. 2004. "Husband-wife disagreement in rural Malawi: A longitudinal analysis." Under Review by Sociological Methods and Research.

[6]. Bignami-Van Assche, Simona, Georges Reniers, Alexander A. Weinreb. 2003. "An Assessment of the KDICP and MDICP Data Quality," Demographic Research S1(2): 31-76.

[7]. Bignami-Van Assche, Simona, Li-Wei Chao, Philip Anglewicz, David Chilongozi, and Agatha Bula. 2007. "The validity of self-reported likelihood of HIV infection among the general population in rural Malawi." *Sexually Transmitted Infections* 83(1): 35-40.

[8]. J Bound, C Brown, N Mathiowetz "Measurement Error in Survey Data". Handbooks In Economics, 2001.

[9]. Miller, Kate, Susan C. Watkins, and Eliya M. Zulu. 2001. "Husband-Wife Survey Responses in Malawi," Studies in Family Planning, 32(2): 161-174.

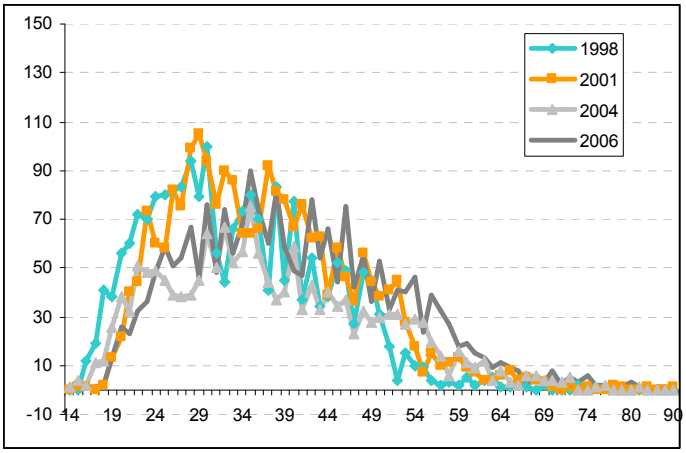
[10]. Cook, K.F., Teal, C.R., Bjorner, J.B., Cella, D., Chang, C.H., Crane, P.K., Gibbons, L.E., Hays, R.D., McHorney, C.A., Ocepek-Welikson, K., Raczek, A.E., Teresi, J.A. & Reeve, B.B. 2007, "IRT health outcomes data analysis project: an overview and summary", Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, vol. 16 Suppl 1, pp. 121-132.

[11]. Hays, R.D. & Lipscomb, J. 2007, "Next steps for use of item response theory in the assessment of health outcomes", *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, vol. 16 Suppl 1, pp. 195-199

[12]. Skrondal, A. & Rabe-Hesketh, S. 2004, *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*, Chapman & Hall/CRC, Boca Raton.

Appendix A

Number if individuals, by age



*figures computed using age reported in each wave