WPS7114

Policy Research Working Paper 7114

Milking the Data: Measuring Income from Milk Production in Extensive Livestock Systems

Experimental Evidence from Niger

Alberto Zezza
Giovanni Federighi
Kalilou Adamou
Pierre Hiernaux

WORLD BANK GROUP
Development Research Group
Poverty and Inequality Team
November 2014

Policy Research Working Paper 7114

Abstract

Milk is an important source of cash and nutrients for many households in developing countries. Yet, the understanding of the role of dairy production in livelihoods and nutritional outcomes is hindered by the lack of decent quality household survey data. Data on milk off-take for human consumption are difficult to collect in household surveys for several reasons that make accurate recall challenging for the respondent (continuous production and seasonality, among others). As a result, the quantification and valuation of milk off-take is particularly difficult in household surveys, introducing possibly severe biases in the computation of full household incomes and farm sales, as well as in the estimation of the contribution of livestock (specifically dairy) production in agricultural value added and the livelihoods of rural households. This paper presents results from a validation exercise implemented in Niger, where alternative survey instruments based on recall methods were administered to randomly selected households and compared with a 12-month system of physical monitoring and recording of milk production. The results of the exercise show that reasonably accurate estimates via recall methods are possible and provide a clear ranking of questionnaire design options that can inform future survey operations.

This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted atazezza@worldbank.org.

Milking the Data: Measuring Income from Milk Production in Extensive Livestock Systems. Experimental Evidence from Niger

Alberto Zezza (World Bank)

Giovanni Federighi (Universitá di Roma 'Tor Vergata')

Kalilou Adamou (ICRISAT, Niamey, Niger)

Pierre Hiernaux (Géosciences Environnement, Toulouse, CNRS)[1]

Keywords: Livestock, Household surveys, Livelihoods, Questionnaire design, Milk, Niger

JEL Codes: C8, Q1

[1] The authors are grateful to all the individuals who took part to the monitoring and surveys in the villages and camps of the Dantiandou district, together with the local authorities, for their patience and active participation to the year round monitoring and the repeated surveys. They are grateful to Moussa Kaka, for having visited and vaccinated the dairy cows, and to Oumar Moumouni and Seybou Garba for most ably and carefully conducting the interviews and monitoring of milk production. They would also like to thank Dr Mohamadou Gandah and his colleagues from the ICRISAT administration office for the efficient logistical support they provided to the field team. Alberto Zezza is indebted to Prospere Backiny-Yetna, Derek Baker, Isabelle Baltenweck, Gero Carletto, Celine Dutilly, Mathieu Lesnoff, Nancy McCarthy, Jane Poole, Steve Staal, Diane Steele, and Matthew Turner for guidance, discussion, and comments at various stages of this project. This work was supported by the Livestock Data Innovation in Africa Project, funded by the Bill and Melinda Gates Foundation.

1. Introduction and background

According to a recent review of the role of milk and dairy product in human nutrition by the Food and Agriculture Organization of the United Nations, “growing consumption of dairy and other livestock products is bringing important nutritional benefits to large segments of the population of developing countries” (Muehlhoff et al., 2013: p. 5). Milk production also offers an important source of cash income to many of the over 200 million poor livestock keepers estimated to reside in developing regions (Thornton et al., 2002; Pica-Ciamarra et al., 2011).

Milk is a good source of dietary fat, energy, protein and other nutrients (Wijesinha-Bettoni and Burlingame, 2013). In particular, milk can provide substantial amounts of nutrients such as calcium, magnesium, selenium, zinc, riboflavin, vitamin B12 and pantothenic acid (Weaver et al., 2013). Milk can help provide children of age 6-24 months that are not being breastfed adequate quantities of fat, which is crucial in their diets because it contains essential fatty acids, facilitates the absorption of fat soluble vitamins, and enhances dietary energy density and sensory qualities (Dewey, 2005).

While the evidence is not unequivocal and more studies are needed, particularly for pre-school children, several intervention and observational studies conclude that milk and dairy product consumption are associated with positive effects on linear growth in children, the effects appearing stronger in children with existing under-nutrition. Milk consumption has also been associated with secular growth in height whether in industrialized and developing countries (Japan, India) or in pastoral societies (Weaver et al., 2013; Hoppe et al., 2006).

A recent review identified only three randomized trials in developing countries that supplemented children with milk and compared outcomes with a nonintervention control group. The review concludes that “both height and weight growth were improved, although in Kenya height was increased only in younger schoolers who were stunted at baseline” (Dror and Allen, 2014).

Another study by the same authors, based on a thorough review of the available evidence, laments that despite the observed increase in milk production and consumption world-wide, child undernutrition and micronutrient deficiencies that could be alleviated by increased intake of milk and other animal source foods remain highly prevalent. In developing countries, both milk and meat intake improve growth indicators, micronutrient status, and cognitive performance (Dror and Allen, 2011).

Milk consumption, like consumption of other animal source foods, may however be expensive for the poor in developing countries, and markets for it poorly developed (McLeod, 2013). Direct access to milk and dairy production can therefore become important for improving nutrition and alleviating poverty as it can provide households with an affordable way to diversify diets, as well as an opportunity to increase their income (Randolph et al., 2007; Smith et al. 2013). Livestock kept can produce a regular supply of nutrient-rich animal source food, providing not only a critical supplement but also diversity to staple plant-based diets (Murphy and Allen, 2003).

Focusing on programming and interventions, Iannotti (2013) found that the available evidence points towards the potential for milk programming to stimulate local production and simultaneously address malnutrition and poverty, in Africa as well as elsewhere in developing countries. For pastoral communities milk is often the sole source of calories and key nutrients, and a major source of cash income (Sadler et al, 2009). Some livestock by-products such as milk and eggs can help the poorer households to mitigate the effects of often large seasonal fluctuations in grain availability (Wilson et al., 2005).

Rawlins et al. (2014) underscore the role of animal donation programs in improving diets and child nutritional status among poor rural families in Rwanda. However, the impacts remain difficult to estimate precisely in such small observational studies. Hoddinott et al. (2013) using Ethiopian data found empirical evidence to support the hypothesis that cow ownership in underdeveloped rural settings is a key driver of the milk consumption and linear growth of young children.

In general, it is hard to appreciate the role of milk and dairy production in household level livelihood studies in developing regions, because of the generally poor state of agricultural statistics in these countries, and because of the practical difficulties in measuring milk production in household surveys. The discussion that follows focuses on Africa, as this is the region where the empirical work of this paper was conducted, but the arguments made apply more broadly.

Despite the importance of the agricultural sector and its critical role in meeting the MDGs, serious weaknesses in agricultural statistics persist. Of the 44 countries in Sub-Saharan Africa rated by the Food and Agriculture Organization, only two are considered to have high standards in data collection while standards in 21 countries remain low (Carletto, 2009). The scope of coverage and completeness also varies widely (see for example, the four-country case studies by Kelly and Donovan, 2008). Knowledge about agriculture and its impact on welfare and equity is limited by the lack of available, high quality, and consistent data on rural households.

Against this backdrop, the statistics on the livestock sector stand out as an area in particular need for improvement. The neglect of the sector by policy makers and researchers is both a cause and a consequence of the current state of affairs with statistical information on livestock. Household-level data and studies on the role of milk production for human nutrition and livelihoods are severely hampered by the difficulty of producing reliable estimates of milk production in small-scale livestock production systems.

The lack of high quality data on the dairy sector hinders both advocacy and policy analysis efforts aimed at informing actions to support livestock-based livelihoods.

Besides the institutional and political neglect, there are important technical reasons that explain why livestock data are particularly scarce or of dubious quality. Collecting data on some major aspects of livestock activities is inherently difficult, because of peculiarities in the production and marketing processes, in the management of livestock assets, and also in the mobility of some population groups that are especially reliant on livestock for their livelihoods (e.g. pastoralists) , that pose particular challenges to data collection.

Milk off-take data are difficult to collect in household surveys because: (a) Lactating females can be milked daily (often twice, mornings and evenings), but with seasonal patterns; (b) Milk varies depending on the lactation stage; (c) Milk can be left to feed young sucklings; (d) Reproductive and lactating females may be present but not necessarily being milked. For these reasons, the quantification and valuation of milk off-take is particularly difficult in household surveys, introducing possibly severe biases in the computation of full household incomes and farm sales.

This paper presents results from a validation exercise implemented in Niger, where two alternative survey instruments were administered to randomly selected households, and then compared with the results of a physical monitoring of milk off-take over a 12 month period. The immediate objective of this work is to draw lessons for questionnaire design by selecting the best performing options and identifying outstanding issues. The ultimate goal is to contribute to a better understanding of the role of animal production in livelihoods and nutrition, which can facilitate more effective policy and program design.

The focus in the paper is on one specific family of household surveys, the Living Standard Measurement Study (LSMS). This is one prominent type of household survey widely implemented in developing countries to monitor and analyze poverty and livelihoods. While this is just one example of a multi-topic household survey for livelihood analysis, we maintain the lesson for questionnaire design assessed with this exercise can be applied beyond LSMS surveys. The typical 'Livestock products' module in LSMS surveys simply asks questions on the quantity of milk off-take produced over a given period, and the quantity and value of sales. Specialized livestock surveys sometimes attempt to obtain information on milk production or off-take at different times in the reproductive cycle of the animal, and on the duration of the lactating period. Yet, other surveys collect milk off-take recordings on a regular basis and correct them with adjustment factors for suckling frequency.

The paper is organized as follows. The next section outlines the overall design of the validation exercise and the survey instruments being tested. Section 3 describes the data, and section 4 presents the results. The concluding section discusses the implications of this work for future data collection, and elaborates on ongoing next steps in furthering this line of work.

2. Testing alternative survey instruments

2.1 The context: Survey validation work in developing countries

In their primer on methods for testing and evaluating survey questions, Presser et al. (2004a, p: 109) note how "pretesting's universally acknowledged importance has been honored more in the breach than in the practice". Even in countries with well-oiled and well-financed statistical systems, pretests are often limited to a rehearsal of survey interviews, usually on a fairly limited number of cases, which are then qualitatively evaluated by the survey teams so as to draw lessons on questions that seemed to pose problems to interviewers or respondents. Sometimes, this is complemented by a quantitative analysis of response frequencies and other simple statistics from the data collected during a pilot survey.

In most cases there is little that is systematic about these tests, despite the existence of techniques geared towards assessing the performance of survey instruments (see e.g. those reviewed in Presser et al., 2004b, and Iarossi, 2006), and very little documentation is provided to users of the data on the contents of such tests. The evaluation of what 'works' is mostly left to the judgment and experience of the survey team.

Increasingly, however, survey practitioners are paying attention to pre-tests as means towards improving data quality. Also, specific methods are being developed, tested and codified and increasingly applied in survey practice. The interested reader is referred to Presser et al. (2004b) for an excellent review of methods such as cognitive interviews, behavior coding, response latency, vignette analysis, experiments, and statistical modeling. While the use of such methods, and their documentation, is more commonly found in OECD country surveys, their application is gaining grounds in low income countries, including in Africa.

Despite the fact that the quality of the data should be of interest to researchers as much as the quantity, it is surprising how little attention the formal validation of household survey data collection has received in the literature. Researchers' preoccupation with data quality results mostly in efforts to design and supervise survey work as well as possible, but very infrequently are the results of such efforts formally tested. There are some notable exceptions however, and our study aims to contribute to this small but growing strand of methodological literature.

Most of the existing literature on survey experiments and survey validation refers to the measurement of household consumption. Beegle et al. (2012) test eight alternative methods of measuring household expenditure, comparing personal diary as the benchmark to other diary and recall formats. They find significant differences between resulting consumption measures, with the correlation between under-reporting and both illiteracy and urban households' status being particularly evident. In addition, Gibson et al. (2013) use data from the same survey experiment in Tanzania to obtain evidence on the nature of measurement errors, concluding that, as expected, errors have a negative correlation with the true value of consumption.

In the context of household consumption, another issue that has been analyzed is the extent to which the length of the lists of consumption items affect estimates of household expenditures. In a study in El Salvador, Joliffe (2001) shows that a more detailed consumption list, results in higher estimates of mean household expenditures (by around 30 percent). This finding has clear implications for the resulting poverty estimates.

The impact of the level of detail of the questionnaire on key indicators has also been investigated in the field of labor market statistics. Dillon et al. (2012) consider if this aspect, together with the type of respondent, can explain the existing widespread variation in measurement of child labor statistics.

Scott and Amenuvegbe (1990) conduct an experimental study on 135 households in Ghana. Each of them was interviewed 11 times at varying time intervals, asking to report expenditure on the 13 most frequently purchased items. In this study, each additional day of recall returns in a 3% decline of the reported daily expenditure.

The choice of reference period is also likely to have considerable impact in several domains. Beegle et al. (2011) test for recall bias in agricultural data, submitting questionnaires with different length of time between harvest and interviews for three African countries. An assessment of whether and how modalities of data collection in agricultural production may affect results is also provided by Carletto et al. (2012).

Using data from two microenterprise surveys in Sri Lanka, De Mel et al. (2009) find that firms under-reported revenues by about 30%, and that requesting them to maintain account books had significant impacts of on both the revenues and expenses they reported, but not on profits. More generally, they argu that questions on profits give truer measures than asking about revenues and expenses.

What this literature shows is how data collection methods matter as much as analytical tools and statistical techniques for the conclusions of a study. Yet, researchers are often ill equipped for judging the extent to which data quality can be affecting their results, whether using data collected by others or data collected as part of their own research, as the survey instruments employed rarely undergo this type of systematic validation. In particular, we are not aware of similar work done for livestock questionnaire design in the context of household surveys in low income countries, which is the reason that motivated a joint effort by FAO, ILRI, and the World Bank (as part of the LDIA and LSMS-ISA projects[2]) to start the survey validation work that is documented in this paper.

2.2 *Milk production recall methods*

As mentioned in the introductory section, LSMS surveys have typically lumped the collection of data on livestock products in one table listing the different products on the rows and a set of standard questions, common to all products and based on a 12-month recall period, on the columns. The module usually asks a variation on two rather simple questions: (1) "Number of production months in the last 12 months", and (2) "Average production per month during production months". Sometimes these questions are asked for milk as a homogeneous product, sometimes the product is broken down by livestock species (cow, ewe, goat) or by dairy product (fresh or curd milk, cheese, butter).

Because of the peculiarities of milk production recalled earlier (continuous production, seasonality, varying lactating capacity of animals over time among them), such simple recall questions are likely subject to large errors, most likely in the way of underestimating annual production. This has led livestock researchers and livestock survey specialists to devise more complex strategies to collect more accurate milk off-take data as well as an expanding set of additional information useful to evaluate milk production systems.

Examples of these elaborate approaches include the 12 months method developed by researchers in France's CIRAD (see Lesnoff et al., 2010), which relies on the monitoring/recording of production over extended periods of time, as well as on techniques which while based on recall approaches, try and prompt the respondent more in depth about the milk off-take system hoping that this will help increase the accuracy of the responses. In developing new survey approaches to be integrated in LSMS-type surveys that include an expanded agricultural focus, these approaches are useful, but need to be adapted to conform to both the objective of the survey as well as to the survey operations. The only way to assess whether a change in approach results in an actual improvement in data quality is to validate the new method via fieldwork, ideally in an experimental setting, while reproducing as closely as possible real survey conditions.

[2] For information on the two projects see http://www.fao.org/ag/againfo/programmes/en/Livestock_Data_Innovation_in_Africa.html and www.worldbank.org/lsms-isa.

It is beyond the scope of the LSMS, in terms of both objective and logistics, the possibility to collect milk production data over extensive time periods, or in a way that allows calculating the complex milk productivity parameters often required by livestock sector specialists. The main goal of LSMS-type surveys is to generate information on household living standards and livelihoods, in this case jointly with information on the productivity, profitability and returns to different activities households may be engaged in. The LSMS survey logistics are organized with mobile teams, that normally reside in each enumeration area for 3-4 days, and need to complete the survey operations in that location in that given time. It is therefore beyond the scope of the LSMS, in terms of both objective and logistics, the possibility to collect milk production data over extensive time periods, or in a way that allows calculating the complex milk productivity parameters often required by livestock sector specialists. The objective of an LSMS needs to be more modest, and limited to collecting a reliable measure of milk production that can accurately portray the role that milk production has in the overall household livelihood strategy.

At the same time, LSMS-type surveys aim to look at the heterogeneity across households, so methods that rely on the application of technical production factors from the literature (e.g. average milk production per animal in a certain environment) to variables that may be easier to measure in a survey (such a number of animals milked by the household) may result in accurate 'average' estimates, but may artificially reduce the observed differences in milk production (both in physical and value terms) across households. For most of the analysis performed with LSMS data, the analysis of the dispersion of the distribution is often as if not more important than the analysis of the measures of central tendency (means, medians). Also, the number of lactating cows, ewes, goats milked, the volume of milk extracted, amount of time milking is practiced for are all management decisions that vary across households and herders, for reasons that include but go beyond the milk production potential of the animal as expressed by technical parameters. For these reasons, competing data collection methods will need to be evaluated not only on the basis of their ability to yield an accurate point estimate of, say, mean milk off-take, but also on their ability to return a distribution of observations that resembles as much as possible the 'true' distribution .

In view of these considerations, in developing the Niger survey validation we looked at two methods that are often applied in livestock sector surveys, but also seemed to hold promise of being adaptable to both the questionnaire design and logistics of LSMS survey operations. In what follows we will refer to these two methods as the "Average Milk per Day" (AMD) and the "Lactation Curve" (LC) methods.

The two questionnaires are identical, except for one question on milk off-take. Both questionnaires are asked at the level of each animal species (cows, ewes, goats, camels), and start off by prompting the respondents about the number of months during which animals were milked for human consumption, and how many animals were milked on average during each of those months. The questionnaires then differ in that the AMD asks for the average quantity per day off-taken during the reference period[3], whereas the LC questionnaire asks about the amount of milk milked on average from each animal at three, or four, different points in time: one week, one month, three and six months after parturition. The two modules then continue asking the same set of questions on issues of whether calves/lambs/kids were allowed to suckle, about the time duration between parturitions, and about the placement of milk off-take (family consumption and sales either fresh or after transformation into dairy products).

Annual household milk off-take production can be calculated from both questionnaires. In the AMD this is done by simply multiplying the average daily production by 30 days (to get to monthly production per animal), then by the number of production months and by the number of animals milked. In the LC methods things are a little more complicated, and annual production is calculated as the area under each animal's lactation curve, or rather milk off-take curve. This is not immediately intuitive and requires some further explanation.

All mammals have a pattern in the lactating period with lactation starting shortly after parturition, a peak in lactation reached early in the lactation period, and then a slow decline in lactation to the end of the lactation period. The timing of these periods, and the overall length of the lactation vary by animal species, and by breeds, and with climatic, grazing, watering and a host of other factors. Besides that, what the survey measures is not lactation as such, but the amount of milk that is taken off for human consumption, which is a decision variable for the farmer.

Total milk off-take can therefore be approximated, assuming a constant value of off-take between the last point in time for which recall is asked and that of the end of the milking

[3] In fact, there are two variants of the AMD method, one that asks about Average Milk *per Animal* per Day (AMAD), and one that asks about Average Milk *per Herd* per Day (AMHD). Our results refer mainly to the latter even though – as we explain in what follows – we used both at different points in our fieldwork.

period[4], as the area under a curve such as the one depicted in Figure 1. In the most general case of four monitoring points, the corresponding formula can be written as:

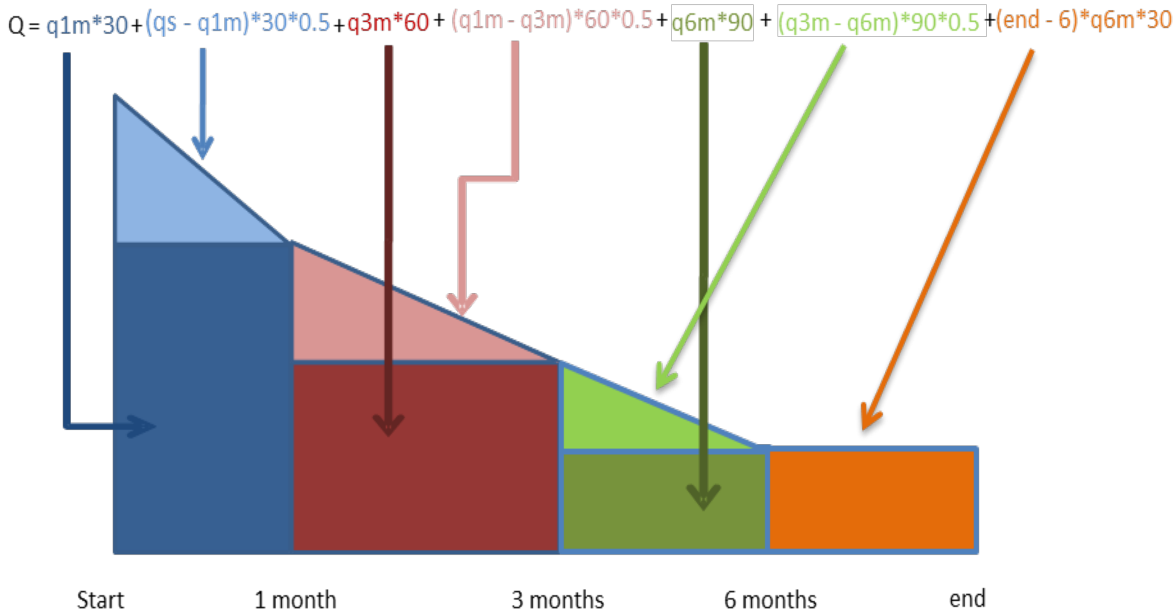
$$Q= q1m*30+(qs-q1m)*30*0.5+q3m*60+(q1m-q3m)*60*0.5+q6m*90+(q3m-q6m)*90*0.5+q6m*(end-6)*30$$

Where *Q* is the total milk off-take per animal in one lactation, *qs* is the average daily quantity of milk off-taken per animal at the start of the lactating period (one week after parturition in the Niger LC module), *q1m, q3m* and *q6m* are respectively the off-take one month, three and six months after parturition, and *end* is the average number of months of milk off-take per animal. For animals with shorter lactation periods such as ewes, and goats, more parturitions (and hence lactating periods) may fall within the 12 months of the survey reference period. In such cases, the presence of a question on the average interval between parturitions allows attributing a quota of the second lactation to the survey reference period (Njiuki et al., 2011). In this paper we focus on cattle milk over a 12 months reference period, which rules out the possibility of multiple lactations for any one animal as the calving interval for cattle is longer than 12 months.

With the LC method respondents are asked to recall more information (milk off-take at different stages of lactation) but to only average out this information across the animals they have milked. In the AMD method, respondents are required to report only one figure, but to obtain that via an implicit process of averaging not only across animals but also across lactation stages. What process is easier for the respondent and more likely to return an estimate closer to the 'true' value is an empirical question, and the main question this papers aims to address. Whether it is easier for respondents to respond to questions about an average animal or about the entire herd is also an empirical question. In the study area each animal is milked separately but the milk extracted is poured in a single pot (or a series of pots), thus the herder in charge of milking may have a feeling for both the average volume of milk from a cow and the average volume of milk collected from the herd. After some piloting in the field, it was felt that respondents found it easier to report about production per herd, as the milk is collected for all animals into one container, once or twice a day.

[4] An alternative way of computing milk production is to assume a monotonic decline in lactation from a peak after a week from parturition, to zero at the end of the lactation (milk production) period.

Figure 1: *Computing milk off-take using the LC method*



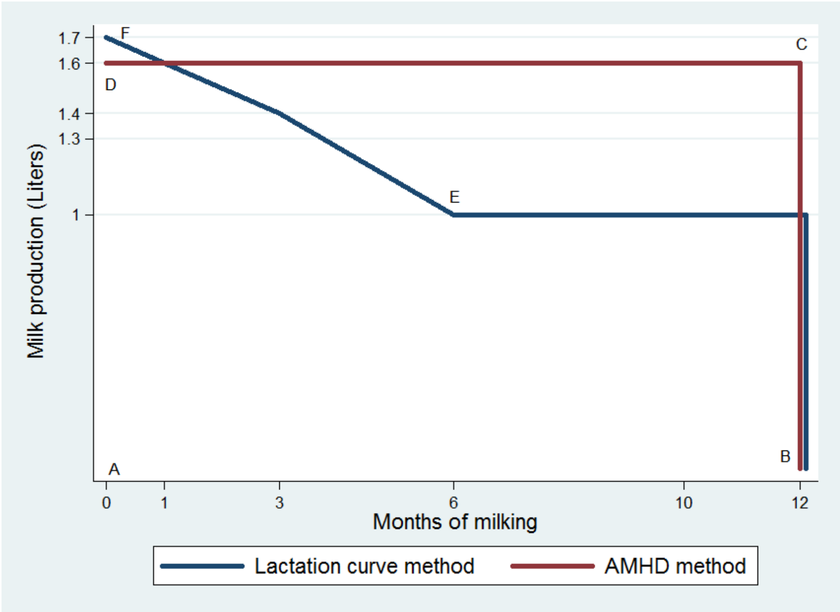
Some livestock survey practitioners suggest that the response given to the AMD question may result in an overestimate of the quantity of milk collected as the response patterns may lead to estimating the area under a rectangle that will largely be above the lactation curve triangle. Figure 2 illustrates the point, using hypothetical values not too dissimilar from the data in our Niger cattle milk off-take study. In calculating total milk off-take from the AMD method one is essentially computing the area of the rectangle ABCD, where AB is the number of months milk was collected and BC is the monthly quantity (in liters) collected milk[5]. Suppose the true shape of the off-take curve for the respondent was equal to the polygon ABEF and it becomes evident how AMD would result in an overestimate of milk production.

The AMD method can be administered for different recall periods, as it is often argued that shorter recall can improve data quality. This is especially true for variables that are characterized by seasonal patterns, which is the case for milk production. In the case of the LC method this is not feasible as a, say, six-month recall period would likely be shorter than

[5] Note that the area of the rectangle ABCD depends on the mean milk off-take by animal by day (BC) but also of the mean duration of milking (AB). The latter is much more variable and depends on factors such as individual, parity, season, and it could reflect a progressive reduction of milking frequency.

the lactation/off-take period, thus complicating the task of the respondent as some of the points in the off-take curve may fall outside of the recall period.

Figure 2: *Comparison of recall methods*



It is also often found that additional questions related to the main object of the recall can be useful in aiding the recall by the respondent. For that reason, in the exercise described in this paper, we also experimented in combining the AMD with the LC questions. The idea is that if a respondent is first invited to recall average milk off-take at different stages over the lactating period, he may then be able to provide a more accurate answer when asked to estimate the average off-take of his herd (or per animal) than if asked to provide that figure directly.

In the exercise reported on here we compare the following methods: (a) the LC method; (b) the AMHD method with a 12 month recall; (c) the AMHD method with a 12 month recall and linked to the LC method questions; (d) the AMHD method with a 6 month recall. All are compared against a benchmark constructed by the physical monitoring of daily milk off-take measured every fortnight over a 12 month period. We also provide some evidence on the performance of the AMAD variant of the AMD method. Before discussing the results of these comparisons, we now turn to describing how the data used in the paper were generated.

3. Data

The main data set analyzed in this paper comes from fieldwork that took place in the Dantiandou district in Niger, between April 2012 and June 2013, and is referred to here as the Dantlait survey. The fieldwork was managed by two experienced enumerators, and a supervisor, all three ICRISAT staff. The team monitored the milk off-take of 300 families over 12 months, as well as associated livestock management, together with family consumption and sale of dairy products. The team also administered 6 month recall questionnaires on 200 families, and 12 months recall questionnaires to 400 families.

The first 200 family farms were randomly sampled among the 835 family farms documented in 2009 and 2010 for the Livestock Climate and Society (ECliS) project (final report and documents on http://eclis.get.obs-mip.fr/). These 835 families live in 13 villages and associated camps within the district (commune and canton) of Dantiandou (80 km East of Niamey). A large data base is available on the composition of the family, its economic activities (including cropping, breeding livestock, forestry, and off farm), the composition and number of livestock, milking practice, consumption and sale of dairy products. This data base was used to stratify the families based on the type of dwelling (either village or camp), which largely matches with socio-ethnic affiliation (Zarma/Fulani), and on the size of the herd (less than 5, 5 to 15, more than 15 adult females). The additional 100 families were selected in additional 13 villages from the district of Dantiandou (5,340 families and 45 villages in total), based on the 2008 national census.

The monitoring method targeted the assessment of the daily milk off-take in each of the 300 sampled families. For each sampled family herd, the milk off-take was measured one day every fortnight adding morning and evening milking when applicable. At each milking, the total milk off-take of the herd was poured in a transparent plastic pot devoted to that measure. The level reached by the milk was marked on the outside of the pot with a marker by the herder. To assess milk volume, the research assistant weighted the plastic pots empty and when filled with water up to the mark done on the side of the pot. The pot weights were recorded on the herd recording form together with the number of lactating females, and the number of lactating females milked re-actualized at each visit. Equipped with a motorbike, each of the two enumerators monitored about ten farms per day (one or two visits depending on milking practices), with revisits every two weeks.

Camp families involved in dairy farming are endowed larger cattle herd in average than village families (7.2 vs 4.4), both however are managing quite small herds. The mean number of lactating cows in the course of the year is 3.4 vs 1.8. Then, only a fraction of the lactating females are actually milked, in average 1.9 vs 1.3. Resulting mean milk off-take are low, at 2.1 liters per day in camps and 1.3 liters per day in village farms. There are large seasonal variations, the wet season and first part of the dry season ('cool' season) contrasting with the late dry season, with milk yield in a factor 2 in camp farms and factor 1.5 in village farms. These seasonal variations are explained by the reproductive cycle of the cows (peak of birth in early wet season), the better quality of grazing resources, but these reasons are mediated by the herder decision (i.e. share of the lactating cows actually milked, milking in the morning/evening or both, volume off-take). It appears for example that the volume milked (0.8 to 0.9 liters per cow and per milking) does not vary with farm type, morning or evening milking, position along the lactating curve. Sparing milk for the calves drives the practice of milk off-take especially in camp farms.

Recall questionnaires were asked to 200 farms (141 of which had also been monitored) in December 2012, and to 400 farms (269 of which had also been monitored) in May-June 2013. The December survey included a 6 month recall AMHD questionnaire. The 400 households interviewed for the 2013 survey were randomly split into two groups, with one being administered a 12 month AMHD recall, and the other a LC module, where the LC questions where followed by an AMHD question. We are therefore able to compute recall measures based on the four measures described above (6 month recall, which we also annualize by multiplying it by 2), LC curve, 12 month AMHD, and 12 month AMHD cum LC recall aid.

The objective of the physical monitoring was to construct a measure that could be used as a benchmark against which the different recall methods are compared[6]. Earlier in the project, a LC questionnaire and a 12-month AMAD recall had been included in the national ECVMA survey implemented in 2011 by the 'Institut National de la Statistique' (INS) of Niger, with technical assistance from the World Bank and the 'Ministère de l'Elevage', on a nationally representative sample of 3,968 households, of which 2,430 are rural and 1,538 urban. While

---

[6] One source of measurement error we do not focus on here, is the tendency of respondents to recall question to report figures in round numbers (e.g. quarter of a litre or half a litre), whereas for physical monitoring measures were recorded in actual cubic centimetres. This is a well know phenomenon known (see Roberts and Brewer, 2001). For an application to land size measures in agricultural surveys see Carletto et al., 2013.

it was not possible to construct a benchmark for this large nationally representative survey, the results of the comparison between the two recall methods can be interpreted in the light of the conclusions emerging from the Dantiandou survey and monitoring.

While the standard LSMS-type livestock product module was not used in these surveys (national ECVMA and Dantiandou monitoring surveys), a smaller scale pilot survey that was run in February 2011 on about 60 households provides qualitative confirmation of the expectation that the standard LSMS module tends to understate milk production compared to other recall methods. As the ECVMA national sample, the pilot survey asked the milk production quantities not only for the entire herd but also for each different livestock species (ewes, goats, cows and camels).

Of importance to the design of the study, we observe no significant differences between the two groups in which our sample was randomly split. That provides confidence in that the random design on which the survey is based worked, and that the groups being compared have no systematic difference other than the fact that they have been asked different questions. Table 1 summarizes the descriptive statistics for the key groups in which the sample has been split for the fieldwork and the analysis. On only one variable we do find statistically significant differences (months of milking) between the LC and the AMD questionnaire. If we look at the physically monitored and the not monitored subsamples, there are some meaningful differences in the number of milking months, and in the average number of raised and lactating cows. However, for a further separation of those who responded to the LC or to the AMD questionnaires into the monitored and not monitored groups, there are not shown any statistically significant differences. Most of the comparisons we will base our conclusions on will bear on the 269 monitored households only, so that even if there was a bias in the selection of the households to monitor, it would not affect the comparisons. The non-monitored households were mainly added to the sample to obtain some more statistical power in the comparison of means.

Table 1: *Summary statistics for different randomly selected sub-samples*

		Avg. raised cows	Avg. lactating cows	Months of milking	Avg. cows milked	Length of previous lactation	Gap in last two births	Age of cow at first birth	Number of births	Age of cow	
questionnaire type // unit of measurement		*units*	*units*	*months*	*units*	*months*	*months*	*months*	*units*	*years*	
Lactation Curve Quest.	*Obs.*	172	170	168	168	175	155	168	169	168	
	Mean	5.64	2.73	11.07*	2.02	12.35	22.17	52.23	3.04	10.55	
	Median	4	2	12	2	12	24	60	3	10	
	Std. Dev.	5.09	1.84	2.31	1.19	4.36	5.31	24.46	1.39	3.14	
	Min	1	1	2	1	4	12	4	1	5	
	Max	31	10	12	7	30	36	108	7	19	
Avg. / Herd / Day (AMHD) Quest.	*Obs.*	168	164	163	163	157	154	164	165	164	
	Mean	5.88	2.76	10.50*	2.02	12.61	22.30	52.23	2.98	10.27	
	Median	4	2	12	2	12	24	60	3	10	
	Std. Dev.	5.51	2.08	2.98	1.18	4.85	5.43	23.98	1.45	3.04	
	Min	1	1	1	1	3	12	3	1	5	
	Max	35	12	12	7	28	36	96	9	18	
Lactation Curve Quest.	*Obs.*	135	135	135	135	138	122	133	134	133	***Monitored***
	Mean	5.93	2.82	11.13	2.01	12.46	22.11	51.62	3.05	10.45	
	Median	5	2	12	2	12	24	60	3	10	
	Std. Dev.	5.16	1.90	2.21	1.11	4.36	5.26	24.45	1.37	3.03	
	Min	1	1	2	1	5	12	4	1	5	
	Max	31	10	12	6	30	36	96	7	19	
Avg. / Herd / Day (AMHD) Quest.	*Obs.*	134	134	134	134	127	125	134	134	134	
	Mean	6.25	2.91	10.69	2.05	12.34	21.96	51.77	3.02	10.38	
	Median	4.5	2	12	2	12	23	60	3	10	
	Std. Dev.	5.81	2.18	2.82	1.17	4.63	5.40	24.33	1.48	3.05	
	Min	1	1	1	1	3	12	4	1	6	
	Max	35	12	12	7	28	36	96	9	18	
Lactation Curve Quest.	*Obs.*	37	35	33	33	37	33	35	35	35	***Not monitored***
	Mean	4.57	2.37	10.82	2.03	11.95	22.42	54.57	3.00	10.94	
	Median	3	2	12	2	12	22	60	2	10	
	Std. Dev.	4.73	1.57	2.69	1.47	4.39	5.54	24.71	1.48	3.56	
	Min	1	1	2	1	4	12	5	1	5	
	Max	25	7	12	7	24	36	108	7	17	
Avg. / Herd / Day (AMHD) Quest.	*Obs.*	34	30	29	29	30	29	30	31	30	
	Mean	4.41	2.10	9.62	1.90	13.77	23.76	54.27	2.81	9.80	
	Median	3	2	12	2	12	24	60	2	10	
	Std. Dev.	3.85	1.37	3.58	1.23	5.66	5.42	22.64	1.30	3.00	
	Min	1	1	1	1	4	12	3	1	5	
	Max	14	5	12	5	24	36	96	7	17	

*Source: Authors' calculation based on data collected for the experiment. Significance levels: * 10%; ** 5%; *** 1%.*

4. Results

The expectation going into the exercise was that the LC method could provide an improvement over the AMD, which we expected to overestimate production. The key results from the validation exercise carried out as part of the ICRISAT-led fieldwork in Dantiandou are reported in Table 2.

The first rather surprising result is that the AMD recall methods do in fact perform much better than was expected, and appear to be superior to the LC methods. The deviation of the median values from the median of the milk monitoring is surprisingly close to the value obtained via the physical monitoring with a difference of just 21 liters (or about 3 percent). The deviations for the mean values are somewhat larger but still acceptable at 30 liters (3 percent of the monitoring value for the 6-month recall, up to 6 percent for the other variants).

Secondly, for the LC Method the results are less satisfactory. Deviations from the 'gold standard' represented by the physical monitoring range between 'acceptable' levels at 6 and 10 percent, when median values are considered (for the 4 and 3 points measures, respectively). If one considers deviations from the average value of the monitoring, however, differences increase to 13 percent for the 4-point LC method and 37 percent for the 3-point LC method. In general, the 4-points method appears to perform significantly better, thus justifying the extra question required of the respondent.

Thirdly, the results show that a major feature common to both the AMD and LC methods is how they over-estimate the dispersion around the mean (as measured by the standard deviation), and particularly so for the LC method. Among the AMD variants, the highest standard deviation is 1.4 times the standard deviation of the monitoring. For the LC method the ratio is 1.8.

Within the AMD methods, shortening the recall period to 6 months appears to perform as well as the 12 month recall, without any major improvement in accuracy. In this particular sample the 6 month recall survey did not generate any large extreme value, which happened for the 12 month survey, but it is hard to generalize this result, as it is linked to the performance of a few respondents7. We will however return to this matter when we discuss other measures we use to assess the relative accuracy of the various methods.

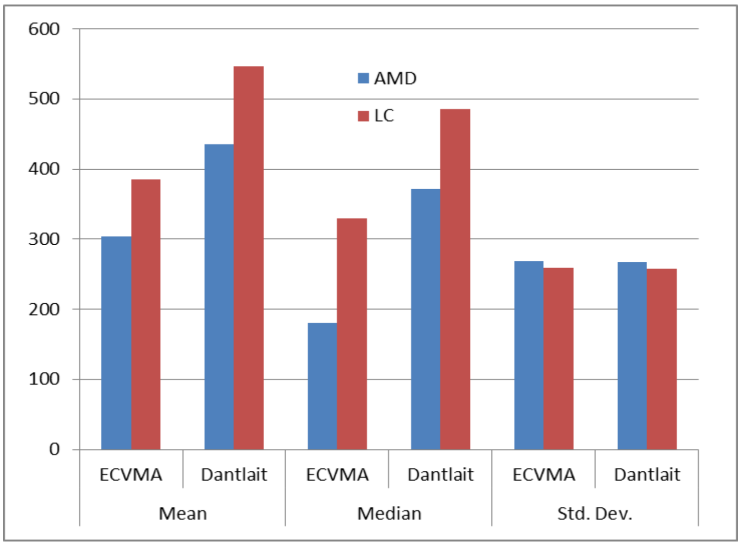
Another result that is interesting to note is how the AMD method, when integrated with the LC questions, returned substantially more accurate results than when the LC questions were not included. It appears, but again this is based on few observations in the left hand tail of the distribution, that introducing LC as a recall aid did help respondents to average out to values closer to the 'true' value, which we approximate here with the monitoring. This is particularly true for camp household, which are characterized by both higher production values, and deeper seasonal fluctuations.

[7] One possible reason is that the 6 month recall survey occurred at the end of the wet season, so that the average milk offtake was mostly based on the high milk off-take during wet season while the 12 month survey occurred at the end of the dry season, requiring for the herder a more difficult averaging exercise between low milk off-take in the last months and higher off-take in the former months.

Importantly, very similar findings regarding the differences between the estimates obtained via the AMD and LC methods are observed in the data collected via the national ECVMA survey, which did not include a benchmark measure as did the Dantlait survey. Figure 3 reports the mean, median and standard deviation measures of milk production per cow in both data sets. The patterns, in terms of differences between the LC (based on 3 points) and AMD methods, are very similar in the two surveys. This is consistent with the idea that the Dantlait survey results can be extrapolated to a sample of households in other parts of Niger, conducted as part of a larger scale, national survey operation conducted by the national statistical office.

It is important to note, however, that the Dantlait survey was limited to cattle. Small ruminants have shorter lactating periods, and the same results may very well not apply to them. In the ECVMA data, for instance, milk off-take from ewes and goats[8] is substantially higher when estimated with the AMD method compared to the 3-point lactation curve method, which is the opposite of what happens for cattle in the same sample. Unfortunately, as discussed earlier, the ECVMA did not include a benchmark that allows assessing the precision of these estimates, and the Dantlait survey only collected data on cattle. Throughout the paper, therefore, we will be referring only to estimates of cattle milk off-take.

Figure 3. *Comparison of mean, median and standard deviation measures of milk off-take estimates from AMD and LC methods in Dantlait and ECVMA surveys (liters)*



Source: Dantlait and ECVMA surveys

[8] Data not reported, but available from the authors upon request.

Besides getting at reasonable average estimates, however, LSMS-type surveys are geared towards depicting the heterogeneity in household's livelihoods and productivity. To that end, looking at how different indicators perform along the entire distribution, and understanding how well they can estimate the position of each household along the distribution is as, if not more, important as obtaining an accurate central tendency measure. For these reasons it is worth analyzing also the correlation and regression coefficients between the different recall methods and the monitoring benchmark (Table 3), and the box plots for the different measures (Figure 4).

Looking at Table 3, it is comforting to observe that the implicit ranking of the different recall methods observed for central tendency (Table 2) is also confirmed when one looks at the overall correlation between the measures resulting from different recall methods. The annualized 6 month (AMD, top row) and the straight 6 month recall (bottom row)9 display the highest coefficients and R2, followed by the other 12 month recall methods in the order in which they appear in the table, and again pointing to a better performance of the 4-point compared to the 3-point LC variant.

[9] The annualized 6 month recall is just the 6 month recall times 2. What really changes in the comparison between the two is the benchmark data, which is the full 12 months of monitoring in the former case, and the first 6 months of monitoring in the latter.

Table 2: *Household milk off-take (liters). Comparison of monitoring and recall data, various methods (annual and 6 months).*

	TOTAL						VILLAGE						CAMP					
	Obs.	Mean	Median	Std. Dev.	Min	Max	Obs.	Mean	Median	Std. Dev.	Min	Max	Obs.	Mean	Median	Std. Dev.	Min	Max
Physical monitoring																		
Monitoring at 12 months	300	877	741	631	10	3291	129	605	512	465	10	2484	171	1083	971	662	45	3291
Recall on 12 months																		
6 months recall - annualized	171	847	720	699	8	3600	63	569	360	534	8	3240	78	1089	1080	640	180	2880
Avg. / Herd / Day (AMHD) - LC module	167	934	720	870	43	5400	55	684	557	591	43	2229	79	1072	929	845	130	4458
Avg. / Herd / Day (AMHD) - All	330	926	720	863	9	6687	111	759	557	692	9	3960	157	1049	743	880	130	6687
Avg. / Herd / Day (AMHD) - 12 months recall	163	918	720	859	9	6687	56	832	549	777	9	3960	78	1027	743	920	130	6687
Lactation curve - 3 points	167	1200	818	1146	132	6900	56	1091	600	1140	132	6037	79	1284	913	1229	201	6900
Lactation curve - 4 points	167	990	693	934	87	6037	56	915	480	1010	87	6037	79	1055	855	954	174	5263
	Note: 330 sample of households with not null milk production.																	
Monitoring at 6 months	300	471	386	323	10	1825	129	334	267	230	10	1321	171	574	509	345	45	1825
Recall at 6 months	171	424	360	350	4	1800	63	284	180	267	4	1620	78	545	540	320	90	1440
	Note: 300 monitored hhs (152 with LC quest. / 148 with AMHD quest.).																	

Source: Authors' calculation based on Dantlait survey data.

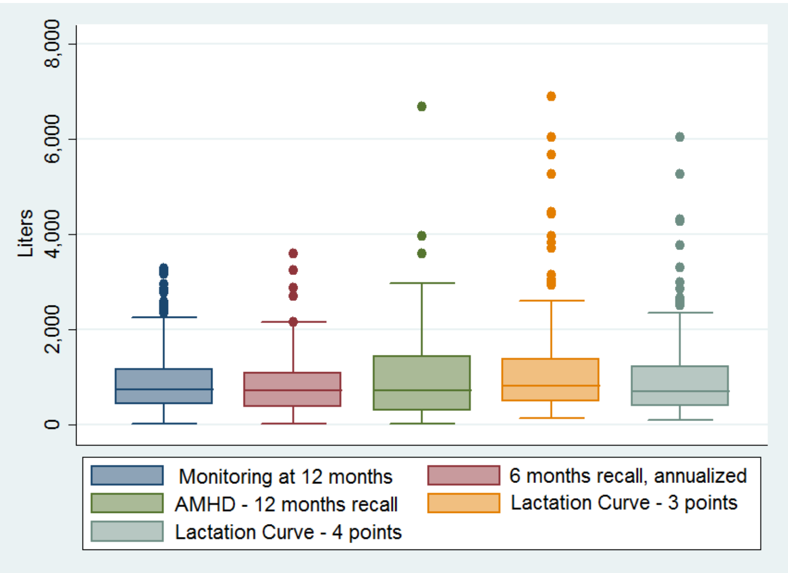
Table 3: Correlation and regression (Ordinary Least Squares, OLS) coefficients between monitoring and recall methods

	Correlation coefficient	OLS no constant		OLS		OLS (logs)		N
Correlation with 12 months monitoring		Coeff	R2	Coeff	R2	Coeff	R2	
6 months recall - annualized	0.71	0.91	0.81	0.68	0.50	0.76	0.63	141
Avg. / Herd / Day (AMHD) - LC module	0.61	0.79	0.72	0.51	0.38	0.57	0.48	134
Avg. / Herd / Day (AMHD) - All	0.52	0.73	0.66	0.41	0.27	0.58	0.44	268
Avg. / Herd / Day (AMHD) - 12 months recall	0.44	0.69	0.60	0.33	0.19	0.58	0.41	134
Lactation curve - 3 points	0.35	0.47	0.52	0.19	0.12	0.47	0.21	135
Lactation curve - 4 points	0.36	0.57	0.53	0.24	0.13	0.49	0.24	135
Correlation with 6 months monitoring								
Recall at 6 months	0.67	0.97	0.78	0.69	0.44	0.76	0.63	141

Source: Authors' calculation based on Dantlait survey data.

The box-plots (Figure 4) provide further support to these results. To improve readability we have only graphed five indicators, the monitoring, 6 month recall annualized, 12 months recall, and 3 or 4-points lactation curve. As in the statistics shown in tables 2 and 3, the 6 month recall method shows a little more dispersion than the monitoring, but in terms of median and overall distribution the fit is overall very good. The dispersion at the top end of the distribution increases with the less precise AMD methods, but remains broadly acceptable (even though it is of course hard and to some extent subjective to define 'acceptable' in this case), and becomes substantially higher for both variants of the lactation curve method.

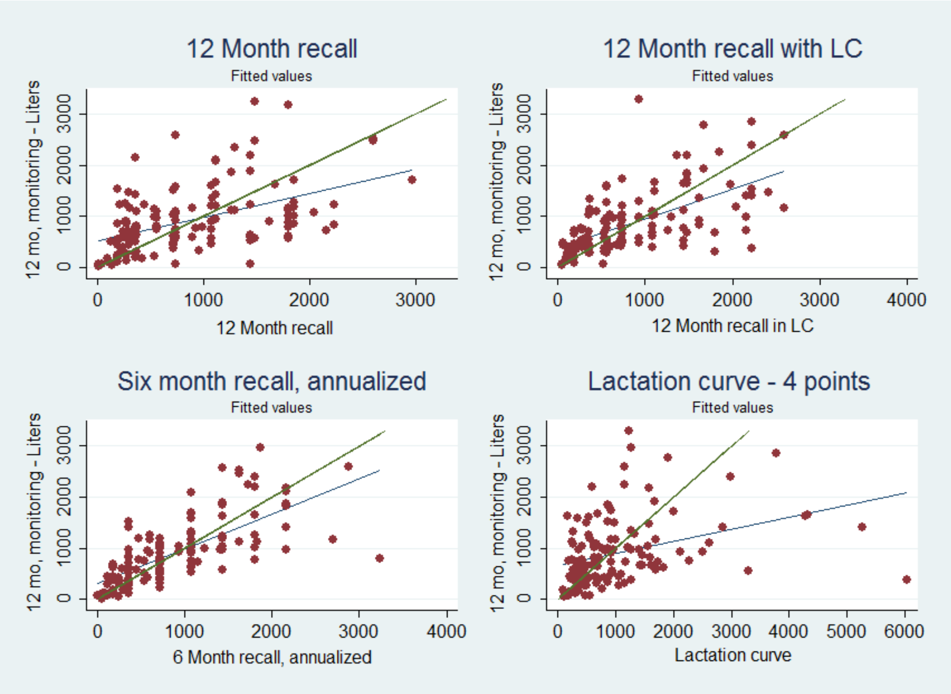
Figure 4: *Box plots of mean household daily milk off-take (liters): Monitoring and recall*



Source: Dantlait survey

Then, to look more closely into the correspondence between the different measures for the individual households, we have plotted scatter plots of the different recall measures against the result of the milk monitoring. Results are reported in Figure 5, where the green line represent the line of equality between the two measures (this would be a 45 degree line if the axes had the same scale), whereas the blue line is based on a linear fit.

Figure 5: *Household milk off-take (liters): Scatter plots of recall against monitoring method*



Source: Dantlait survey

A few things are notable from these graphs. First, the methods that perform better when judged on the synthetic measures we have analyzed so far, also perform better when we look at individual household observations. The cloud is a lot more scattered in the case of the LC method than it is for the six month recall or the 12 month recall with the LC aid. Second, a fair amount of measurement error remains10. More importantly, at this visual inspection the error does not seem to be randomly distributed, but tends to be negatively correlated to 'actual' (i.e. monitored) milk production. Respondents are more likely to under-report milk production if they produce larger quantities of milk, and they are more likely to over-report

[10] It should be noted that while we treat milk monitoring as the benchmark this measure is also, as any measure, affected
by some degree of error.

production when they produce smaller quantities. This is clearly a matter of concern for the analyst, as measures of income from milk production and productivity based on such data would be biased on ways that are correlated with other variables of interest.

For that reason, it is important to understand what are the correlates and determinants of the observed measurement error. Table 4 presents the results of a series of linear regressions where the percentage difference between the recall methods and the monitored milk quantities (the dependent variables, one for each method) are regressed against a set of covariates which we expect to be able to influence the quality of the recall. The independent variables in the regressions include herd and production system characteristics, as well as other household and respondent features.

Since we expect respondents to be less accurate in averaging out production over 12 months the greater the day-to-day variability in production levels, the first variable in the regression is the household-specific coefficient of variation of the monitored production, computed as the standard deviation of total milk produced for all cows divided by its mean. We also include variables that reflect differences in management or milking practices that may be systematically related to recall quality: whether the household is in a village or camp, whether cows are milked once or twice per day, the number of cows milked, and the duration of the last lactation period. A variable measuring the number of cows that receive feed supplements is included, as this indicator can be related to both milk production per cow, as well as managerial ability or availability of resources on the part of the herder.

We hypothesize that respondents that are not exclusively focused on cattle rearing might recall events about livestock less accurately, and we use information on ownership of other animals, engagement in activities other than agriculture and source of income from migration as additional controls. The number of mobile phones owned by the household is included as a proxy for overall wealth, as well as ability to access and process information, while age of the household head is included on the grounds that ability to recall may decline with age. On the other hand, if younger farmers are less experienced, response accuracy could actually increase with age. Since two different enumerators collected data in the field, we also include a dummy to control for possible differences in enumerator’s ability.

The most consistent, robust message that comes from this analysis is that the measurement error is correlated with the number of animals milked. The coefficient is of the expected positive sign, large in magnitude and highly statistically significant in five of the seven regressions we estimated. Interestingly, the coefficient is not significant only in the two methods based on the 6 month recall, suggesting that shortening the recall period may be an effective means to not only improve accuracy but also reduce bias.

Respondents living in the camps appear to be better able to recall the amount of milk production, and this is reflected in smaller measurement error. This may be linked to management practices, to the fact that livestock might be relatively more important in camps, and to reasons of ethnicity (Fulani herders are more likely to be residing in camps, compared to Zarma). It is hard to disentangle these effects and it should also be acknowledged that for reasons of collinearity, one should interpret with caution the regression coefficients that relate to management practices. This is the case for instance for the puzzling positive sign on the coefficient for the dummy capturing whether cattle are milked twice per day. We expected fewer milkings per day to be associated with better recall quality, but in fact it seems to be associated with greater measurement error.

The negative coefficient on the supplementation variable and positive coefficient on the duration of lactation variables, on the other hand, are expected, but only statistically significant for the lactation curve methods. We explain the former as reflecting greater managerial ability or simply greater importance given to animal management, and the latter to be related to the fact that the longer the milking period, the greater the degree of approximation implicit in the estimate of production employing the lactation curve method, and the related formula.

Turning our attention on some other possible household characteristics, we found little or no impact for the other household characteristics, which is not surprising given the relative homogeneity in the socio-economic composition of the villages studied. There are other two factors which we would have wanted to control for, namely the educational level and the gender of household head. Unfortunately, the level of education of the population in the district of Dantiandou is extremely low, even by Nigerien standards, and virtually our entire sample of households is headed by a man. In other settings these variables may however play a role. We take comfort in the result that measurement error does not appear to be influenced by the enumerator collecting the data.

Table 4: Regressions' results on the determinants of the measurement errors.

	(1) 6 months recall annualized	(2) Avg / Herd / Day - LC module	(3) Avg / Herd / Day - All	(4) Avg / Herd / Day - 12 months recall	(5) Lactation Curve 3-points	(6) Lactation Curve 4-points	(7) Recall at 6 months
Coefficient of variation	-0.514 (0.317)	-0.127 (0.551)	0.468 (1.075)	0.463 (1.698)	-0.986 (1.148)	-0.527 (0.960)	-0.253 (0.230)
Dummy Territory (1=Camp 0=Village)	0.121 (0.229)	-0.404* (0.209)	-0.393 (0.323)	-0.233 (0.486)	-0.780** (0.339)	-0.618** (0.279)	0.252 (0.182)
Dummy =1 if milk is collected only in the morning	0.290 (0.186)	0.367* (0.195)	0.680** (0.303)	0.974* (0.574)	0.430 (0.323)	0.405 (0.290)	0.250* (0.137)
Enumerator dummy	-0.119 (0.186)	-0.016 (0.173)	-0.036 (0.234)	-0.108 (0.473)	0.073 (0.376)	-0.107 (0.318)	-0.073 (0.135)
Log number of cows	-0.106 (0.142)	0.677*** (0.190)	0.783*** (0.204)	0.962*** (0.327)	1.232*** (0.380)	1.143*** (0.371)	-0.033 (0.120)
Number of supplemented cows	-0.052 (0.039)	-0.040 (0.050)	-0.046 (0.033)	-0.068 (0.053)	-0.258*** (0.079)	-0.222*** (0.069)	-0.003 (0.032)
Log of age of hh head	-0.095 (0.277)	0.510** (0.242)	0.235 (0.355)	0.126 (0.664)	0.405 (0.514)	0.423 (0.449)	0.141 (0.180)
Annual household remittances received (1,000 CFA)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
Dummy =1 if agriculture is the only hh activity	0.028 (0.151)	-0.391* (0.229)	-0.370* (0.201)	-0.410 (0.347)	0.013 (0.339)	0.048 (0.310)	0.036 (0.126)
Number of mobile phones owned	0.102 (0.063)	0.040 (0.098)	-0.027 (0.077)	-0.192 (0.157)	0.151 (0.195)	0.182 (0.198)	0.060 (0.056)
Number of other animals	0.051 (0.047)	0.021 (0.051)	-0.050 (0.053)	-0.165 (0.112)	0.018 (0.069)	-0.003 (0.062)	0.023 (0.039)
Duration of previous lactation	0.002 (0.015)	-0.017 (0.024)	-0.041 (0.028)	-0.051 (0.041)	0.175*** (0.060)	0.142** (0.057)	-0.006 (0.010)
Constant	0.511 (1.068)	-1.523 (1.187)	-0.232 (1.384)	0.726 (2.225)	-2.418 (2.485)	-2.711 (2.386)	-0.743 (0.689)
Observations	129	134	266	132	135	135	129
Log – likelihood	-146.6	-181.9	-533.9	-299.1	-254.8	-240.2	-115.2
Prob. > F	0.773	0.000	0.001	0.124	0.000	0.000	0.413
R-squared	0.105	0.199	0.091	0.100	0.370	0.340	0.094
Adj. R-squared	0.013	0.120	0.048	0.009	0.308	0.275	-0.000
RMSE	0.795	0.990	1.846	2.457	1.681	1.508	0.623

Note: Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Finally, it is interesting to note how the overall fit of the lactation curve models is much higher compared to the other recall methods, whereas the simple 6 month recall has the lowest (with an adjusted R-squared equal to zero). That suggests that the lactation curve methods likely embed a larger degree of systematic error which correlates with several variables of interest related to livestock management, which is hardly a desirable feature when employing a productivity measurement in analytical work.

5. Conclusions

Agriculture plays a key role in the livelihoods of poor households throughout the developing world, and particularly so in Sub-Saharan Africa. Livestock accounts for an increasing share of food consumption, offering increasing opportunities for value addition and income generation. For many poor households, livestock ownership is a major contributor to satisfying the needs for cash as well as protein. For the households more prone to climatic stresses, such as those in arid areas, it is often the main - if not the only- source of livelihood. Milk in particular offers a marketable product that is available throughout the year and that can play an important role for nutrition, particularly for children.

While there has been a renewed interested in the research over the nexus between agriculture, poverty and nutrition in recent years, associated with the increase in international food prices, this has not been matched by an improvement in the state of agricultural statistics. In Africa the availability and quality of agricultural sector data leave much to be desired, and that is particularly so for the livestock sub-sector. In terms of methods, livestock statistics offer peculiar challenges that are exemplified by the difficulties of collecting accurate milk off-take data at the household level. However, of the small investments in livestock statistics, hardly any goes into methodological validation. The work documented in this paper takes the motivation from this state of affairs, and the belief that given the abysmally low level of attention to this type of work, effort to improve data quality can have substantial marginal returns and multiplier effects on research and policy analysis.

There are some clear messages we take away from work implemented in Niger to test different recall methods to capture household level milk off-take data, against a gold-standard of physical monitoring over a 12-month period.

The first is that even though there is a substantial amount of measurement error in the way even the best recall methods we tested perform in capturing household milk off-take, some methods do in fact perform fairly accurately, and much more accurately than what we expected when we designed this exercise. In particular, the methods do a reasonable job at estimating the more common central tendency measures (mean and median), as well as the distribution of milk production across sample households.

The methods that rank consistently better among those we compared are the 6 month AMD recall, and a 12 month AMD recall coupled with a lactation curve recall aid. The lactation curve method, on the other hand was consistently the worst performer, with differing patterns depending on the number of data points used to estimate the off-take level at different points in the lactation. Within the AMD method, the shorter recall period appears to significantly improve the estimates, even though it is uncertain the extent to which this result would hold if the 6-month recall interview were to be moved to another point in time, given the seasonality of milk production.

While we did find some evidence of the AMD method being more likely to return some extreme values (which is one of the perceived shortcomings of this method), this occurrence was rather limited in our sample and not frequent or large enough to undermine the overall performance of the method. In particular, not only did the AMD methods yield more accurate estimates of average milk production in our sample, but they also provided a more accurate depiction of the 'true' distribution, something that is as important when assessing the role of milk production (and livestock) in general in livelihoods and attempting to capture the heterogeneity across households.

Another reason militating against the use of the lactation curve method and in favor of the AMD, is that the former seems to not only lead to larger measurement error, but also to a greater likelihood of measurement error being correlated to other variables of interest, such as herd size and length of the milking period, and hence of total milk production itself.

Last but not least, the LC method is somewhat more demanding on the respondent (who is prompted a few more questions) as well as the analyst, who needs to derive milk off-take estimates from the calculation of the area under the milk off-take curve as described in equation (1). To achieve the same result, the AMD method requires fewer questions, and a much simpler multiplication of daily average production times the length of the off-take period.

The main limitation of the study concerns its external validity, that is, the extent to which the conclusions that can be made based on our data apply to survey data collection in other areas in Sub-Saharan Africa or in other developing regions, and to animals other than dairy cattle. Both concerns can only be addressed by replicating similar methodological validation exercises in different settings. Ancillary evidence to the results presented in the paper do point to the fact that the distributions of the milk production estimates may perform very differently for large and small ruminants, due to the shorter lactation periods of the latter. But again, this speculation can only be verified through further research.

Taken together, the results presented in this paper have clear implications for future questionnaire design that we feel are strong enough to recommend using the better performing methods in future household surveys of small livestock keepers in extensive livestock systems in low-income settings. While there are limits to the external validity of these results, which should be repeated in different settings and for different species, we do maintain that the findings reported here are strong enough to be already taken up in future questionnaire design by National Statistical Offices, researchers, and anyone involved in household survey design.

REFERENCES

Beegle, K., De Weerdt, J., Friedman, J., Gibson, 2012. "Methods of household consumption measurement through surveys: Experimental results from Tanzania". *Journal of Development Economics* 98 (1): 3-18.

Beegle, K., Carletto, G., Himelein, K., 2011. "Reliability of Recall in Agricultural Data". World Bank Policy Research Working Paper 5671. 5671.

Carletto, G., 2009. "Improving the Availability, Quality and Policy Relevance of Agricultural Data: The Living Standards Measurement Study – Integrated Surveys on Agriculture". Development Research Group, the World Bank, Washington, DC.

Carletto, G., Deininger, K., Savastano, S., Muwonge, J., 2012. "Using diaries to improve crop production statistics: Evidence from Uganda". *Journal of Development Economics*, 98 (1): 42-50.

Carletto, G., Savastano, S., Zezza, A., 2013. "Fact or Artefact: The Impact of Measurement Errors on the Farm Size-Productivity Relationship". *Journal of Development Economics*, 103: 254–261.

De Mel, S., McKenzie, D.J., Woodruff, C., 2009, "Measuring microenterprise profits: Must we ask how the sausage is made?", *Journal of Development Economics*, 88 (1): 19–31.

Dewey, K., 2005. "Guiding Principles for Feeding Non-Breastfed Children 6-24 months of Age". Geneva, World Health Organization.

Dillon, A., Bardasi, E., Beegle, K., Serneels, P., 2012. "Explaining variation in child labor statistics". *Journal of Development Economics*, 98 (1): 136-147.

Dror, D.K., Allen, L.H., 2011. "The importance of milk and other animal-source foods for children in low-income countries". *Food and Nutrition Bulletin*, 32 (3): 227-243.

Dror, D.K., Allen, L.H., 2014. "Dairy product intake in children and adolescents in developed countries: trends, nutritional contribution, and a review of association with health outcomes". *Nutrition Reviews*, 72 (2): 68-81.

Gibson, J., Beegle, K., De Weerdt, J., Friedman, J., 2013. "What Does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?". World Bank Policy Research Working Paper 6372. 6372.

Hoddinott, J., Headey, D., Dereje, M., 2013. "Cows, missing milk markets and nutrition in rural Ethiopia". In: Presentation at the Farm Production and Nutrition Workshop, World Bank, June 7.

Hoppe, C., Mølgaard, C. and Michaelsen, K.F., 2006. "Cow's Milk and Linear Growth in Industrialized and Developing Countries". *Annual Review of Nutrition*, 26: 131-173.

Iannotti, L.L., 2013. "Milk and dairy programmes affecting nutrition". In Muehlhoff, E., Bennett, A., and McMahon, D. (eds.). *Milk and dairy products in human nutrition*. FAO Publications, Rome.

Iarossi, G., 2006. "The Power of Survey Design. A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents". World Bank: Washington, DC.

Joliffe, D., 2001. “Measuring absolute and relative poverty: The sensitivity of estimated household consumption to survey design”. *Journal of Economic and Social Measurement*, 27 (1): 1-23.

Kelly, V.A., Donovan, C., 2008, “Agricultural Statistics in Sub-Saharan Africa: Differences in Institutional Arrangements and their Impacts on Agricultural Statistics Systems. A Synthesis of Four Country Case Studies”. MSU International Development Working Paper No. 95.

Lesnoff, M., Messad. S, Juanes, X., 2010. 12MO. “A cross-sectional retrospective method for estimating livestock demographic parameters in tropical small-holder farming systems”. CIRAD: Montpellier, France.

McLeod, A., 2013. “Human nutrition and dairy development: Trends and issues”. In Muehlhoff, E., Bennett, A., and McMahon, D. (eds.). *Milk and dairy products in human nutrition*. FAO Publications, Rome.

Muehlhoff, E., Bennett, A., and McMahon, D. (eds.), 2013. *Milk and dairy products in human nutrition*. FAO Publications, Rome.

Murphy, S.P., Allen, L.H., 2003, “Nutritional Importance of Animal Source Foods”. The *Journal of Nutrition*, 133: 3932S – 3935S.

Pica-Ciamarra, U., Tasciotti, L., Otte, J., Zezza, A., 2011. “Livestock assets, livestock income and rural households. Cross-country evidence form household surveys”. FAO, Rome.

Presser, S., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J.M., Singer, E, 2004a. “Methods for Testing and Evaluating Survey Questions”. *Public Opinion Quarterly*, 68(1): 109-130.

Presser, S., M.P. Couper, J.T. Lessler, E. Martin, J. Martin, J.M. Rothgeb, E. Singer (eds.). 2004b. *Methods for Testing and Evaluating Survey Questions*. Wiley.

Randolph, T.F., Schelling, E., Grace, D., Nicholson, C.F., Leroy, J.L., Cole, D.C., Demment, M.W., Omore, A., Zinsstag, J., Ruel, M., 2007, “Role of livestock in human nutrition and health for poverty reduction in developing countries”. *Journal of Animal Science*, 85: 2788-2800.

Rawlins, R., Pimkina, S., Barrett, C.B., Pedersen, S., Wydick, B., 2014. “Got milk? The impact of Heifer International’s livestock donation programs in Rwanda on nutritional outcomes”. *Food Policy*, 44: 202-213.

Roberts, John M., Brewer, Devon D., 2001. Measures and tests of heaping in discrete quantitative distributions. *Journal of Applied Statistics*, 28 (7): 887–896.

Sadler, K., Kerven, C., Calo, M., Manske, M., Catley, A., 2009. “Milk Matters: A Literature Review of Pastoralist Nutrition and Programming Responses”. Feinstein International Center, Tufts University and Save the Children, Addis Ababa.

Scott, C., Amenuvegbe, B., 1990. “Effect of Recall Duration on Reporting of Household Expenditures”. World Bank Social Dimensions of Adjustment in Sub-Saharan Africa Working Paper No. 6.

Smith, J., Sones, K., Grace, D., MacMillan, S., Tarawali, S., Herrero, M., 2013. “Beyond milk, meat, and eggs: Role of livestock in food and nutrition security”. *Animal Frontiers*, 3 (1).

Thornton, P. K., Kruska, R. L., Henninger, N., Kristjanson, P. M., Reid, R. S., Atieno, F., Odero, A., Ndegwa, T., 2002. "Mapping poverty and livestock in the developing world". International Livestock Research Institute, Nairobi, Kenya.

Weaver, C., Wijesinha-Bettoni, R., McMahon, D., Spence, L., 2013. "Milk and dairy products as part of the diet". In Muehlhoff, E., Bennett, A., and McMahon, D. (eds.). *Milk and dairy products in human nutrition*. FAO Publications, Rome.

Wijesinha-Bettoni, R., Burlingame, B., 2013. "Milk and dairy product composition". In Muehlhoff, E., Bennett, A., and McMahon, D. (eds.). *Milk and dairy products in human nutrition*. FAO Publications, Rome.

Wilson, T., Pearson A., Bradbear, N., Jayasuriya, A., Laswai, H., Mtenga, L., Richards, S., Smith, R., 2005. "Livestock products – Valuable and more valuable". In: Owen, E. A., Kitalyi, A., Jayasuriya, N., Smith, T. (eds.). "*Livestock and Wealth Creation: Improving the Husbandry of Animals Kept by Resource-Poor People in Developing Countries*". Nottingham University Press, UK.