

Erich Battistin (University of Maryland), Michele De Nadai (University of New South Wales), and Nandini Krishnan (World Bank)

The Insights and Illusions of Consumption Measurement: Evidence from a Large Scale Randomization

We revisit the issue of whether diaries are the most appropriate benchmark at a time when developing countries have increasingly moved to recall. Contrary to several influential studies discussed below, we find that the presumption that diaries yield better data is just an illusion. Diaries have become the workhorse for benchmarking comparisons of household measurements from alternative collection modes. Beegle et al. (2012) randomized households in Tanzania to assisted individual diaries to define a benchmark against alternative survey designs that use recall questions. This is possibly the most important experiment on consumption measurement in low- and middle-income countries in the last decade. Accuracy has been a longstanding concern not only in developing countries. Recall questions on food spending from the Canadian Food Expenditure Survey were compared to expenditure diaries. A similar exercise was done for the Consumer Expenditure Surveys (CEX) in the United States. In these instances, the accuracy of survey reports is assessed against diaries. The idea that assisted diaries on food spending yield better data is widespread.

The assumption that diaries provide an error-free benchmark is done for convenience but has little statistical justification. Diaries are far from perfect. Most surveys are characterized by differences in reporting within the diary period, usually with a declining pattern over time. Possible explanations are declining cooperation due to fatigue, but also the tendency of participants to deviate from the usual purchasing behavior as an effect of the diary or social pressure. We also know that the ratio between food totals from household surveys and national accounts can be considerably lower using diaries than recall questions ([32], and [10]). Moreover, diaries are expensive: regular visits to ensure quality standards and data processing boosts costs.

Data quality can be a red herring in the study of the distributional aspects of welfare using diaries. One key aspect here is the difference between consumption and spending, which has been largely overlooked in empirical work. Household surveys elicit consumption or expenditure on food with the idea that they should yield similar figures. With the increasing shift towards recall, most surveys in developing countries focus on food consumed, the monetary value of consumption, and the sources from which food was acquired. Recall modules ask households to report on food consumption during the specified period, and to value that consumption, irrespective of source, at market prices. Diaries instead measure

household spending for market acquisitions, and market-valued consumption for gifts, self-production and other non-market sources. The comparison between diary and recall data may be misleading if the former source primarily measures spending and the latter consumption. The frequency of purchases may vary across food items, differentially across socio-economic groups. While assumptions can be made to claim that spending and consumption are the same on average, as we discuss below, heterogeneity in the frequency of purchases will typically increase dispersion and has an impact on indicators of poverty and inequality (in the cross-section and over time). Assuming that diaries measure expenditures without errors is not enough to account for these problems.

Differences between experimentally-formed groups of respondents to diary and recall interviews are, in general, only partially informative about survey errors. The assumption that spending and consumption have the same average, combined with the assumption of perfect recording of diaries, identify the average error committed by recall respondents. Unless spending coincides with consumption, differences in welfare indicators (e.g., the Gini coefficient) between the two experimental groups are not solely the result of data quality. This is true also when diary entries are complete, and entries have no errors. For example, even random deviations of diary spending from household consumption would raise the density of diary data where it is convex (e.g., in the tails) and flatten it where it is concave (e.g., at its mode), increasing dispersion.

We leverage a unique large-scale experiment in Iraq designed for the Household and Socio-Economic Survey (IHSES) in 2012. The survey ran continuously, year-long, and was administered to about 25,000 households across the country. All households filled out a 7 day diary on their spending patterns, assisted by enumerators during regular visits. One-third of households in each primary sampling unit were randomized to an additional survey module, administered before the diary, asking them to recall food consumption in the last 7 days. Both modules had the same respondent, the household head or the most informed adult. The size of the “treatment” sample, of roughly 8,000 households, is therefore of a different order of magnitude compared to other studies in developing countries. The household diary employed was standard, with one respondent recording on behalf of the household. The recall module was specifically designed to inform the national statistics agency on the transition from diary to recall, which is planned for 2019-2020. A list of recall expenditure groups was selected based on an assessment of their importance in household food expenditure shares and on how commonly they were reported, based on IHSES diaries for 2007.

Compared to previous studies, our innovation is in the use of differences between diary and recall measurements for the same household, combined with differences in the reports of

households assigned to different experimental groups. We use these multiple measurements to study the quality of reports from diaries and recall modules allowing for survey effects in both measurements. Our model is non-parametrically identified under mild conditions. It yields identification of survey error distributions for recall and diary reports without assuming that errors from alternative collection modes are mutually independent, or independent of true consumption. The distribution of true consumption underlying an household's diary and recall measurements is also non-parametrically identified. We can therefore assess the extent to which diaries deliver mean and distributional indicators close to those that would be computed using true consumption.

The availability of repeated diary-recall measurements of food spending for the same household is not unusual, the CEX Diary Survey and the Canadian Food Expenditure Survey being well-known examples. Our approach therefore extends to a number of contexts beyond the specific case study.