Caloric intake and energy expenditures in India

Shari Eli and Nicholas Li
University of Toronto

Abstract: We estimate total energy expenditure for the Indian population between 1983 and 2012 to shed light on the debate concerning falling measured caloric intake during the period (Deaton and Drèze (2009)). We use anthropometric and time-use data to estimate the separate components of total energy expenditure related to metabolism and physical activity levels, combined with detailed employment surveys. Despite a significant drop in adult physical activity levels we find that total energy expenditures are flat overall between 1983 and 2012. Rising metabolic requirements due to increases in weight dampened the effect of falling activity levels on total energy expenditure. In addition, the 10% decline in the population share of children in the period raised average total energy expenditures considerably as children have much lower metabolic requirements and activity levels than adults.

Acknowledgments: We would like to thank Eric Edmonds, two anonymous referees, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, Ronald Lee, Ted Miguel, Gustavo Bobonis, Stephan Litschig, Daniel Lafave, Pranab Bardhan, Tom Vogl, Anne Case, Jeff Hammer, as well as participants of seminars at Berkeley, Princeton, Toronto, NEUDC and PAA. Lucas Parker provided outstanding research assistance.
1. Introduction

Caloric intake and food expenditures are often viewed as key indicators of an individual’s ability to satisfy the most basic of needs and as good predictors of net nutritional outcomes, which has made these indicators fundamental to the measurement of household welfare since the work of Ernst Engel (1895). Applications include the construction of India’s national poverty measures, guided by household caloric “norms” of 2400 calories per day in rural areas and 2100 in urban areas, as well as equivalence scales (Barten (1964), Deaton and Muellbauer (1986)). In this context, Deaton and Drèze (2009) note the puzzling finding that based on India’s National Sample Survey (NSS), per capita caloric intake in India fell substantially between 1983 and 2005 despite rising per capita expenditures. They consider explanations ranging from measurement error in consumption data to consumer-driven changes (due to tastes, prices, or changes in the set of available goods) to changes in physical activity levels and the disease environment that may have lowered the demand for calories, but reach no firm conclusion.¹

In this paper we estimate Total Energy Expenditures (TEE) for the Indian population over the 1983-2012 period to provide insight into this puzzling finding. Total energy expenditures are the total calories required for maintaining one’s weight level given the body’s metabolic processes and physical activity levels. We use anthropometric data from three waves of India’s National Family Health Survey (1998, 2005, 2015) to estimate Resting Energy Expenditures (REE) that capture metabolic requirements related to height, weight, age and gender for individuals at rest. We use detailed time-use data from India’s 1998-1999 Time-Use Survey, matched to activity factors in FAO/WHO/UNU Expert Consultation (2001), to estimate activity levels (AL) that capture the additional energy used

¹Note that the “puzzle” documented by Deaton and Drèze (2009) for India has also been documented in other contexts, e.g. Du et al. (2002) for China and Clark et al. (1995) for Great Britain during the Industrial Revolution. We discuss literature on measurement error and the disease environment in the Indian context below. Other contributions include Basole and Basu (2015), who focus on the role of prices and food budgets squeezed by rising expenditures on fuel, education and medical goods, Patnaik (2010), who argues that higher relative prices for food lowered demand, and Gupta (2013), who argues that conspicuous consumption (Veblen goods) could explain lower food intake.
for physically intensive activities. We then use the large set of variables in common between these surveys and the NSS employment surveys (Schedule 10) that were conducted between 1983 and 2012 to estimate individual TEE. This quantification provides insight into the evolution of population TEE and the contributions of changes in demographics, occupation and industry, work status, and labor-intensive domestic tasks.

Our main finding is that population level TEE in India is fairly flat between 1983 and 2012 and actually increased slightly over the 1983 to 2005 period before declining slightly between 2005 and 2012. Our quantification highlights two reasons for flat TEE over time. First, individual and household characteristics that predict lower Activity Levels (AL) also tend to predict higher Resting Energy Expenditures (REE) for a given demographic composition. For example, higher levels of education, higher levels of expenditure, and less labor-intensive occupations are all associated with lower AL but higher REE. Because TEE is computed as the product of REE and AL (FAO (2001)) as we discuss later, the effect of changes in education, expenditures, and occupation on TEE tend to be smaller than their effect on AL. Overall, AL fell and REE increased over this period controlling for demographic composition. Our analysis also uncovers evidence of household scale economies in TEE. Households with more members, but similar demographic composition and per capita expenditures, have a lower AL that is only partly offset by higher REE. Although household scale economies contribute little to our main findings on population TEE, this supports the conjecture in Deaton and Paxson (1998) that “caloric overheads” (fixed costs of common household activities in developing countries, such as travel for work, buying/selling goods, or collection of water or fuel) could explain the common empirical finding that larger households have lower per capita food expenditures holding constant demographic composition and per capita expenditures.

Second, our estimates imply that TEE fell by over 100 calories per day over this period for individuals aged 15 or older, which on its own would be enough to account for most of the decline in caloric intake in NSS data during this period. However, over the
1983-2012 period, India underwent substantial demographic change, with the share of the population aged 14 and under falling from 40% to 30%. Children have much (about 40%) lower TEE than adults, due to both lower REE (due to smaller size) and lower AL (particularly in rural areas where primary/middle-school attendance appears much less physically active than adult activities). Thus even though TEE is falling for both adults and (to a lesser extent) children throughout this period, demographic change at the population level is more than enough to offset this “within” effect, except after 2005 when activity levels fell particularly quickly.

We conclude with a discussion of the potential contribution of measurement error and changes in the disease environment to the Deaton and Drèze (2009) puzzle of declining measured caloric intake in light of our estimates of flat population TEE and modest gains in nutritional status over time. Smith (2015) argues that the NSS estimates of caloric intake may be biased down, especially in recent years, due to rising consumption of food away from home.\(^2\) We review the evidence on the magnitude of recall bias identified using different survey recall periods (National Sample Survey Organization (2000), NSSO (2013)) and also the sensitivity of estimated caloric intake to assumptions about the caloric value of food expenditures. We also discuss our findings in the context of Duh and Spears (2017), who test one of the potential mechanisms discussed in Deaton and Drèze (2009). Duh and Spears (2017) show that infant mortality and other indicators of disease related to nutritional absorption (e.g. latrine ownership, diarrhea, and open defecation) are strongly associated with within-district changes in NSS caloric intake between 1988 and 2005 as well as in the cross-section. They find that this pattern holds even when controlling for household and individual variables that capture much of the variation in TEE explored in this paper, concluding that an improved disease environment can account for a substantial decline in caloric intake in India because households “lose” less calories to disease and therefore choose to consume less. Relative to their paper, our contribution is

\(^2\)See also Fiedler and Yadav (2017) for evidence from the 2011-2012 NSS survey and Smith et al. (2014) for a general review of the reliability of food data collected in consumption and expenditure surveys.
to test a different potential mechanism discussed in Deaton and Drèze (2009), using the most detailed available data on physical activity and physical size to directly quantify TEE. Overall, our finding that TEE is fairly flat over the 1983-2012 period, combined with the evidence on modest gains in nutritional status, suggests that most of the decline in NSS caloric intake is a result of measurement error and lower caloric burden of disease rather than large decreases in population TEE.

Our paper is organized as follows: Section 2 describes the data and methodology we use for estimating total energy expenditure and discusses some important sources of variation across households, Section 3 presents our results for changes in total energy expenditure over time and discusses the broader context of changes in caloric intake measured in the NSS, and Section 4 offers some concluding comments and suggestions for future research on TEE.

2. Data and measurement of components of Total Energy Expenditure (TEE)

We first briefly outline the methodology and data used to measure TEE from household datasets. The Data Appendix contains a more detailed description including robustness to alternative assumptions. TEE measures the amount of energy used by the human body during a given period. When TEE equals caloric intake there is no weight gain or loss. To calculate TEE, we use the factorial method described in FAO (2001). Indian Council of Medical Research (2009) uses this method to estimate prescriptive caloric requirements for Indian households, based on the 95th percentile height and weight observed in rural areas and providing recommended daily caloric intakes for sedentary, light, moderate, and heavy physical activity. We use the factorial method to generate descriptive estimates of population-level total energy expenditure using the height, weight, demography, and activity-levels we observe in the Indian population using various data sources. The fac-
torial method computes TEE for individual $i$ using the following equation:

$$TEE_i = REE_i \times AL_i$$

where $i$ is an individual, $REE_i$ is the Resting Energy Expenditure of individual $i$ and $AL_i$ represents the Activity Level of individual $i$. REE makes up the bulk of TEE (about two thirds for the average person) and reflects the energy required for essential metabolic functions. An individual in a complete state of rest has $AL = 1$, $TEE = REE$, and still needs a lot of calories to avoid weight loss. An individual that does more than rest will have an AL greater than one. The factorial method implies that a given reduction in AL generates a larger reduction in TEE for individuals with larger REE. Thus, when estimating levels or changes in TEE for a population, one cannot simply average $AL$ or $REE$ across individuals but must sum up the individual $TEE$ as constructed above. We now turn to the measurement of these two components for the Indian population.

2A. Resting Energy Expenditures (REE)

Although REE can only be directly measured in a laboratory setting, there are numerous published predictive equations that are determined by regressing measured REE on more commonly observed attributes such as age, sex, height and weight. While these equations miss an important idiosyncratic element of metabolism, they typically yield an $R^2$ over 0.7. We use the equations provided in Henry (2005) which are based on regressing REE on height and weight separately by sex and age groups (1-3, 4-9, 10-17, 18-30, 31-60, 60+) for a large population of individuals drawn from around the world including tropical and developing countries.

The anthropometric micro-data we use come from India’s National Family Health Sur-
vey (NFHS), a household survey that measures height and weight for a subset of the Indian population. Eligibility for anthropometric measurement within sampled households varies by year – in 1998, children under 3 and ever-married women aged 15-49 were measured, while in 2005 and 2015 all children under 5, all women aged 15-49, and random subsample of men aged 15-54 were measured. We report the results of applying the Henry (2005) equation in the first panel of Table 1.\(^4\) As discussed in the Appendix, we also use data generated by the National Nutritional Monitoring Bureau (reported in Indian Council of Medical Research (2009)) on height and weight by age for rural areas of 16 Indian states in 2000-2002 to extrapolate outside of the age range in the NFHS. The NFHS estimates shown in Table 1 highlight three facts that will be important later: (1) adult men have higher REE than adult women who have higher REE than children, (2) REE is higher in urban than rural areas as urban individuals are slightly taller and substantially heavier for a given age/sex, and (3) REE has been increasing over time for a given age and sex, mostly due to increasing weight (see Online Appendix Table 2).

2B. Activity Levels (AL)

We estimate activity levels by aggregating up from the most detailed activity data available. If we have data on the share of time allocated across \(J\) different activities, we can estimate individual \(i\)’s activity level (AL) as follows:

\[
AL_i = \sum_{j=1}^{J} Share_{ij} \times AF_j
\]

where \(Share_{ij}\) is the fraction of individual \(i\)’s time spent on activity \(j\). \(AF_j\) is an activity factor for activity \(j\) that measures the intensity of activity \(j\) relative to rest (equal to one). Individual \(AL\) are thus weighted averages of activity factors where the weights vary.

\(^4\)Validation studies by Shetty et al. (1986) and Ferro-Luzzi et al. (1997) find that these types of prediction equations, which are typically estimated on European descended populations, have a reasonably good fit for subjects in India.
The most detailed activity data for India comes from the Time-Use Survey (TUS) conducted in 1998-1999 by the National Sample Survey Organization for six states (Gujarat, Haryana, Madhya Pradesh, Meghalaya, Orissa, and Tamil Nadu) that make up 25% of India’s population and are close to representative in terms of regional variation. These data cover all individuals aged 6 or older from the sampled households. Individuals report their activities, classified into 154 activities, in 20 minute increments based on a 24 hour recall period, with variant days capturing weekends/market days. FAO (2001) reports activity factors (derived from field measurement) for a long list of activities that are common in developing countries. We match each activity in the TUS to an activity factor, with some adjustments to deal with codes that have no obvious match (e.g. “other” categories and travel) discussed in the Appendix.

Panel B of Table 1 reports our estimates of activity levels separately by rural and urban residence, sex, and two age groups (6-14, 15 and over) for individuals in the TUS. The data indicate that rural activity levels are substantially higher than urban levels, with the average rural-urban gap equal to about 75% of the population standard deviation for adults. The rural-urban gap is considerably smaller for children, and children in general have much lower activity levels than adults, reflecting the relatively low activity levels associated with school attendance. Although we could only find two studies that directly estimate activity levels in India, by directly measuring REE and TEE for tiny samples of “free-living” adults, our estimates are in line with those reported in Borgonha et al. (2000), who find an AL of 1.9 for rural men near Bangalore, and Krishnaveni et al. (2009), who find an average AL of 1.45 for boys and girls aged 8 to 9 in Mysore.

2C. Imputation of TEE to NSS data
As REE and AL cannot be directly estimated for the same data set using the factorial method, and the coverage in terms of age, state, and time period of the NFHS and TUS is
incomplete, our estimates of TEE will use NSS data that is representative at the population level. To do this, we take advantage of the large number of common variables between the NFHS, TUS, and NSS Schedule 10 Employment Survey (hereafter NSS-E). Essentially, we estimate predictive equations for REE or AL by regressing them on a set of individual and household variables that are common to NSS-E, and then using these predictive equations to generate $\hat{TEE}_i = \hat{REE}_i \times \hat{AL}_i$ for each individual. We focus on the “thick” NSS-E survey rounds beginning with the oldest available (1983) through to the most recently available (2011-2012).

We leave a discussion of the details of this procedure for the Appendix, including the full set of variables included and the estimated coefficients, but depending on the data set (NFHS or TUS) these variables include individual and household demographics, education levels, 2-digit occupation and industry codes at the household and/or individual level, work status (school, domestic work, casual versus salary versus self-employment), labor-intensive “domestic” tasks typically carried out by females not in the labor force (e.g. collecting firewood or water, husking paddy and grinding grains), labor-intensive agricultural tasks, land possessed, and total expenditures (deflated using an all-India CPI). When restricting estimation to the six states in the TUS, we also include state fixed effects. The $R^2$ from these predictive equations ranges from 0.14 to 0.35 for the NFHS REE estimates and from 0.37 to 0.63 for the TUS AL estimates, suggesting the variables in common with the NSS-E are rich enough to capture a lot of the sample variation in REE and AL from the other data sets. We note that while our estimation procedure is adequate for estimation population or conditional means, which is the goal of our paper, it would need to be amended if we were interested in the variance or other moments of the distribution (e.g. fraction of households with TEE below some threshold).

2D. Analysis of household characteristics

While Table 1 makes it clear that demographic composition and rural/urban location will
generate large differences in TEE (and its REE and AL components) across households, there is also considerable cross-sectional variation related to other variables. Before turning to our estimates of population TEE over time, we highlight a few interesting patterns in the cross-section that demonstrate the richness of our data. For this analysis, we use the 1998-1999 Time-Use Survey and impute REE using the procedure described above.

Figure 1 plots household TEE (averaged across individuals within the household) against mean years of schooling for members aged 18 and up, monthly per capita expenditures, and household size. Panel A shows that TEE declines in years of schooling, a net effect that combines a sharper decline in AL and a partly offsetting rise in REE. Panel B shows a similar pattern for per capita expenditures, except that in this AL rises with per capita expenditure for the poorest households, likely reflecting a combination of unemployment/underemployment for the poorest households and a higher proportion of non-earning children. TEE is falling in household size, particularly in the range of 2 to 6 members, which is mostly driven by lower AL, although the unconditional relationship in Figure 1 partly reflects demographic composition.

Our finding that TEE is falling in household size is notable given the observation by Deaton and Paxson (1998) that for several developing countries, measured food expenditure per capita declines in household size. A similar pattern holds for measured caloric intake. Deaton and Paxson (1998) argue that this is a “puzzle” in the sense that larger households with a given demographic composition should benefit from economies of scale that increase their per capita consumption of relatively private/rival goods like food. They suggest that “caloric overheads” – energy expenditures that only need to be incurred by one household member on behalf of the entire household – could be one possible explanation. The TUS data is rich enough to capture differences in time spent on domestic chores, travel, etc. that would be hard to detect in most household surveys.

To examine this further and to also control for demographic composition, Table 2 estimates a log-log regression specification similar to the one used by Deaton and Paxson
The dependent variable is log household average TEE, REE or AL. We include detailed demographic controls for age/sex ratios, log household size, log per capita expenditures, and mean years of adult education. When including all household sizes (1-14), we find a close to zero partial elasticity of TEE to household size, resulting from a perfect offset of REE and AL. Larger households do economize on energy, but their members are also physically “larger” on average. When we restrict to household sizes between 2 and 8, which make up 92% of the sample, we find that there is a significant negative partial elasticity of about 0.017, due to a negative effect on AL that is much larger than the positive effect on REE. Both expenditure and education reduce $TEE$ holding constant demographics and appear to have independent effects.

Columns 7 and 8 of Table 2 estimate the same specification but replace TEE with caloric intake estimated for the 1993-94 NSS 50th round Schedule 1 (consumption survey) in the same states. The coefficient is similar in magnitude (-0.029 using households sized 2 to 8) suggesting that scale economies in energy expenditures can potentially explain much of the apparent decline in caloric intake with household size.\(^5\) Note that the coefficient on education has a similar negative coefficient for caloric intake as for TEE, but per capita expenditures only has a large positive coefficient for caloric intake.

The patterns we document here in the cross-section foreshadow one of our main findings in the time-series, which is that REE often moves in the opposite direction of AL in terms of covariance with observable household characteristics. This dampens the effect of changes in these characteristics on changes in TEE. Online Appendix Figure 1 shows that this pattern also occurs across household main industry – households headed by agricultural or construction workers have the highest activity levels in India but also have the

---

\(^5\)We say apparent because Gibson (2002) and Gibson and Kim (2007) show that much of the puzzling decline first documented in Deaton and Paxson (1998) is reduced when using survey instruments less subject to recall errors, implying that a substantial part of the measured decline in most household surveys is simply measurement error that is negatively correlated with household size. However, they still find some “real” decline in caloric intake with household size when using the alternative survey instruments. See also Beegle et al. (2012) and Brzozowski et al. (2017) for more evidence on food consumption survey measurement error correlated with household size.

3A. Population estimates

Table 3 presents our main estimates of population average TEE for India between 1983 and 2012. We report three separate panels for all India, rural India only, and for the six states in the TUS only. The first column reports the survey-weighted population average TEE (expressed in calories per day) based on our estimates. We find that TEE is fairly flat over the entire 1983-2012 period, rising by only 0.8%. There is a small increase over the 1988-1994 and 2000-2005 periods, and a small decrease from 2005-2012, but these changes are small relative to plausible measurement and estimation errors. Columns 2 presents the population average REE which rose by 6% throughout the 1983-2012 period. Column 3 presents the population average AL which fell throughout the 1983-2012 period by about 4%, mostly after 2005.

Recall that the product of population average REE and AL is not population average TEE, and that changes in population AL and REE capture both changes within age/sex demographic groups and changes in the demographic composition of the population. We report the estimates of population TEE and AL for males 15 and over (columns 4-5), females 15 and over (columns 6-7), and children aged 14 and below (columns 8-9). TEE declined substantially for adult men (135 calories) and women (86) calories over the 1983-2012 period but by less for children (30). Activity levels for adult men and women fell over 6%, and these decreases were only partly offset by their rising REE. For the 1998-1999 TUS, the decrease in AL is equivalent to about one third of the urban-rural gap for adult men and one half for women.

Although we view the decline in AL we estimate for adult men and women as plausible and quite large, it is instructive to consider what variables in the NSS-E drive these results. We report descriptive statistics for all years using the NSS employment survey
in Online Appendix Table 4. Agriculture is one of the more physically active industries/occupations and declined considerably as a share of employment, from 60% to 43% as a share of main household industry; however, there was large and partly offsetting increase in construction work, from 3% to 12%, which is also very physically active. Within agriculture, the probability that individuals engage in specific manual labor intensive agricultural task in the previous week has only fallen slightly, e.g. 12% to 9% for harvesting activities, 2% to 1% for plowing. The share of women engaged mostly in domestic duties remains very high. Although some of these duties are very physically active, most have declined only slightly, with the largest effect coming from the decline in the probability of bringing water from outside the household premises from 11% to 5%, smaller effects coming from decreases in food processing and animal husbandry, and no change in the probability of collecting firewood or cooking/heating fuel (around 7% throughout the period). Our predictions use a more flexible specification, but the linear estimates from Table 2 are also instructive. The 3.3 year increase in years of schooling for adults and the over 50% increase in real per capita expenditure imply only a modest decrease in AL over this period (about 3%) for demographically similar households, with a small offset from the decrease in average household size from 6.3 to 5.3 individuals.

The fact that TEE is falling for adult men, adult women and children yet rising for the population reflects the important role played by population-level demographic change. The share of children aged 14 and under in the population fell from about 40% in 1983 to 30% by 2012. This implies considerably higher TEE, not just because children are physically smaller and have lower REE, but because children are also much less active than adults according to the TUS estimates, particularly in rural areas. In 1983 child TEE was about 60% of adult TEE, implying that the demographic shift over the 1983-2012 period, holding constant average adult and average child TEE, would generate a 4.7% increase in population TEE.
3B. Comparison of TEE with caloric intake, measurement error, and estimates of disease effects

Table 4 combines several data sources to provide an overall assessment of trends and levels for nutrition for India between 1983 and 2012. Columns 1 and 2 provide an illustration of the basic puzzle raised by Deaton and Drèze (2009). Column 1 shows that real per capita expenditure (deflated to 1993-1994 rupees) has increased, particularly in the most recent years since 2005. Column 2 reports official estimates of caloric intake from the National Sample Survey Organization (NSSO (2013)) that are almost identical to the numbers reported in Deaton and Drèze (2009) for the 1983-2005 period but incorporate the latest available data (NSS Rounds 66 and 68 conducted in 2009-2010 and 2011-2012). The official numbers show that caloric intake declined further from 2005 to 2010, but then increased considerably from 2010 to 2012, which erased almost half of the decline between 1983 and 2005 measured in Deaton and Drèze (2009).

Columns 3 and 4 of Table 4 provide some insight into measurement error related to recall bias. It is well known that longer recall periods can lead to downward biased estimates of consumption, and the 51st through 54th NSS rounds (1994-1999) randomly allocated a 7-day as well as traditional 30-day recall period for food across households. The results confirmed that consumption estimates are higher for the 7-day recall period, but also that the difference was much larger for food categories like fruits and vegetables, eggs, fish and meat, and beverages and processed foods (over 40%) compared to under 15% for categories like milk and cereals (National Sample Survey Organization (2000)). These findings led the NSS to use two different recall periods (canvassed for different households) for the food categories with the highest discrepancy between 7-day and 30-day recall periods in 2010 and 2012. For these categories, the magnitude of 7-day/30-day bias was similar to the earlier 1994-1999 period at around 44% for caloric intake ( NSSO (2013) Tables 5A and 5B).

Note also that the use of 30-day and 7-day recall period for the same households in the 55th NSS round (1999-2000) is likely to account for some of the unusual pattern in Table 4 whereby caloric intake rose from
These facts about recall bias in the NSS have two implications for our analysis. First, assuming that the 7-day recall period is more accurate, the overall level of caloric intake in the NSS data using a 30-day recall period is substantially biased downward. The difference in levels of caloric between the two recall periods in 2010-2012 is as large as the entire measured decline between 1983-2005 or 1983-2012 using the consistent 30-day recall period, and this underestimates the overall bias since categories like cereals and milk (which in 2010 and 2012 are measured using a 30-day recall period for all households) also exhibit a downward bias for 30-day recall, albeit to a lesser degree (National Sample Survey Organization (2000)). A higher estimate of caloric intake is more in line with our TEE estimates, particularly in recent years, when 30-day measures of intake are well below our population TEE estimates. Second, the change in caloric intake over time using the 30-day recall period likely overstates the decline in caloric intake that would have been measured using 7-day recall, because the share of expenditures on the goods with the highest degree of bias rose from 36% to 50% between 1983 and 2010. We provide an illustrative estimate of what caloric intake in earlier rounds would have looked like under the assumption that the degree of bias for the 7-day recall categories in 2010 and 2012 remains constant over time. This assumption seems reasonable going back to 1994 in light of the results in National Sample Survey Organization (2000) but we stress that our estimates are intended to be illustrative and need to be treated with caution, particularly for the 1980s where there is no direct measurement of 7-day vs. 30-day recall bias. Column 3 reports the official results using the 7-day recall period for select foods in 2010 and 2012 (from NSSO (2013)) and our imputations going back to 1983 that adjust caloric intake upwards but by a smaller percentage for earlier years (reflecting the change in the expenditure share of these categories reported in column 4). The result of this exercise is that the decline in caloric intake is about 20% smaller over the 1983-2005 period considered in 1994-2000 and fell from 2000-2005. The discrepancy between 30-day and 7-day recall quantities appear smaller when the recall periods are implemented simultaneously for the same households, but the level of 30-day recall is likely to be higher than for households that are not subjected to an additional 7-day recall period (see Deaton and Kozel (2005) for a discussion).
Deaton and Drèze (2009) or the entire 1983-2012 period.

Beyond the potential effects of recall periods on measurement error, numerous imputations are required to estimate caloric intake from a given NSS survey and these imputations have become more important over time. Our independent estimates of caloric intake using NSS data, reported in column 5, provide a sense of the magnitudes of this source of measurement error. We discuss the details of our procedure in the Appendix, but our main departure from the standard NSS practice is how we impute calories for most expenditures in the beverage and processed food categories (which includes purchased "meals"). These expenditures rise from about 5% to 9% of average household food expenditure, and as they lack quantities or report units like "number of cooked meals" or "number of cold beverages: bottled/canned," untested assumptions about how to impute their caloric value have become more important. Our procedure yields very similar levels of estimated intake for 1983 through 1994, but a substantially smaller decline between 1994 and 2005, such that the total decline between 1983-2005 is only 66% as large as estimated by NSSO (2013) or Deaton and Drèze (2009). Our estimates begin to diverge even more after 2005 – while we find a similar pattern of decline from 2005-2010 and recovery from 2010-2012, overall we find a decline between 2005 and 2012 of about 40 calories per day as opposed to an increase of 47 per day according to official estimates. Our point is not that our assumptions are better than the official procedure, as there is no "ground truth" against which to assess either measure. Rather, the point is that measurement assumptions can have a large effect even using the same data, particularly in recent years when the importance of cereals and other unprocessed agricultural staples has fallen. These imputation assumptions have the potential to generate bias of a similar magnitude as the recall errors that potentially bias down both the level and decline in caloric intake.7

Note that per capita caloric intake levels under any methodology are somewhat sensitive to outliers, and we follow the standard practice of censoring the distribution around the 1% and 99% percentile cutoffs (corresponding to 960 and 4400 calories a day per capita). However the trend in population averages over time are not very sensitive to outliers and generally track the population median.

7
Returning to the other side of the calorie equation, we reproduce our estimates of population TEE and average TEE for adults. Relative to the official estimates using the 30-day recall period, our estimated population TEE is almost perfectly negatively correlated over different time periods – it is rising from 1983-1994, stable over 1994-2000 when intake appears to increase, peaks in 2005 when caloric intake is near its minimum, and then declines from 2005-2012 when caloric intake rises. Furthermore, by 1994 intake using 30-day recall is below TEE and by 2005 it is much lower. The levels of TEE we estimate in later periods are more consistent with the estimated caloric intake using the 7-day recall period. Our independent estimates of caloric intake (using the 30-day recall) are also lower than TEE in the later period, but display a similar decline between 2005-2012, such that TEE could “explain” about half of the decline in our caloric intake estimates during this period. In column 7, we report average TEE for adults to once again underscore that, were it not for demographic change, changes in estimated TEE for adults would be enough to “explain” a decrease in caloric intake of 34 calories per day between 1983-2005 (about 20% of the 150 calorie decline in the official estimates) or 111 calorie per day between 1983-2012 (over 100% of the 102 calorie decline in the official estimates).

Further evidence that changes in TEE may have contributed to lower caloric intake between 2005-2012 but are unlikely to have done so for the 1983-2005 period comes from an analysis of state-level changes. Figure 2 plots the change in state caloric intake (using our estimates) against the change in state TEE for the 1983-2005 period and the 2005-2012 period. In the earlier period, although there are some states that appear to line up along the 45 degree line, most lie well below, indicating that measured caloric intake relative to TEE declined. The association is also very weak and not statistically significant. In the later period, the relationship is considerably stronger, with most states lining up along the 45 degree line and a statistically significant slope of 0.666 (with associated $R^2$ of 0.08). Using the full state panel we find that our TEE estimates are significantly correlated with
changes in caloric intake within states over time.\textsuperscript{8}

Given that we find rising population TEE during the 1983-2005 period of falling caloric intake considered by Deaton and Drèze (2009), we also report the best estimates of the potential effect of the disease environment on caloric intake from Duh and Spears (2017). They use a district panel to show that within-district decreases in infant mortality between 1988 and 2005 are associated with within-district decreases in NSS caloric intake, with a slope coefficient of about 1.7. They find smaller but still highly significant cross-sectional associations between village/block infant mortality and caloric intake using a different survey, with slope coefficients ranging from 0.426 (when including household and demographic controls) to 1. They also report results based on open defecation and latrine ownership, but we focus on infant mortality as the data are available for the entire period and plausibly capture changes in the disease environment related to diarrhea, parasites, and intestinal distress that could impede nutrient absorption and potentially raise TEE. Column 9 of Table 4 reports the level of infant mortality (per 1,000 live births), which declined from 106 to 42 over this period. This can be interpreted as a mid-level estimate of the total caloric burden of disease using the slope coefficient of 1 and linear specification reported in Duh and Spears (2017), and we also report a low (column 8) and high (column 10) estimate based on coefficients reported in their paper. Quantitatively, the decline in infant mortality mirrors the fall in caloric intake between 1983 and 2005, although the timing of the decline does not fit as well with the between-round variation (e.g. the huge drop in caloric intake between 2000 and 2005). If these estimates capture the true caloric burden of disease, they also imply that caloric intake measures (official, or our own) using the 30-day recall period must almost surely be too low, particularly in recent years, but the estimated caloric burden of disease is of the correct magnitude to reconcile our TEE estimates with the official 7-day recall period estimates.

A final piece of evidence on net nutrition comes from anthropometric data. Deaton

\textsuperscript{8}See Online Appendix Table 7 for these results.
and Drèze (2009) recognize that there has been some increase in average height in India during the 1983-2005 period, a trend that has continued since although the level of height and the rate of increase remain low compared to other countries with similar GDP per capita and economic growth trajectories. Smith (2015) cites additional evidence of improving nutritional outcomes for children. A rise in average body mass for individuals of a given age and sex is one of the factors contributing to the increase in population REE we estimate in this paper. Although changes in height undoubtedly reflect nutrition, the quantitative relationship between net caloric surplus (intake minus TEE and any caloric cost of disease) and height is not well known. Changes in population height may be more reflective of micro-nutrients, macro-nutrients composition, and the timing of life-cycle nutrition from conception to adolescence rather than changes in the average caloric surplus of the population. However, the NFHS allows us to measure weight gain over time, which does have a well known quantitative relationship with caloric surplus. While a single cross-section only allows us to compare weight across cohorts, the availability of multiple NFHS rounds allows us to estimate average weight gain with age for adults within cohort. In Online Appendix Table 8, we estimate the average weight gain controlling for cohort (birth-year) fixed effects across survey rounds, which means average weight gain between 2005 and 2015 for men and women, and between 1998 and 2005 for ever married women. We also report estimates controlling for height, which may capture some non-random attrition within a cohort related to nutritional status. We find an average weight gain of between 0.5KG and 0.6KG per year within cohort, with a slightly higher rate between 2005-2015 than 1998-2005 for ever married women. A weight gain of 1 kilogram is associated with a caloric surplus of 7700 for adults (St. Jeor and Stumbo (1999)), so the reported level of weight gain translates into between 10.5 ($= 0.5 \times 7700/365$) and 12.7 ($= 0.6 \times 7700/365$) surplus calories per day above TEE and the caloric cost of disease.

---

9In Online Appendix Table 6, we find even larger increases in REE for adults and children and hence smaller decreases (or even small increases) in TEE over the 1983-2012 period when we include a linear time trend in our REE estimation. This may be reasonable in later period but is likely to be a heroic assumption when extrapolating a linear trend observed between 2005 to 2015 back 22 years.
ease. This suggests that, on average, the Indian (adult) population has been in caloric surplus since at least 1998 and even more between 2005 and 2015, even as official NSS caloric intake estimates using 30-day recall plummeted during the 2000s to reach levels well below our TEE estimates. Given that our quantification suggests a limited scope for such a large decrease in population TEE, and that most of the improvement in the disease environment occurred before this period, we interpret the overall evidence as consistent with the view espoused in Smith (2015) that the steep decline in official caloric intake estimates beginning around 2000 is likely the result of substantial measurement error.

4. Conclusion

In this paper, we quantified population level Total Energy Expenditures (TEE) for India over the 1983-2012 period using a combination of time-use data, anthropometric data, and the NSS Schedule 10 employment survey. Our main finding is that population TEE was fairly flat over this period, rising slightly until 2005 and then declining slightly only in the 2005-2012 period. This finding is driven by two main features of the data. First, household characteristics that tend to predict lower activity levels, such as education, expenditures, and industry/occupation, also tend to be associated with greater height and weight that imply offsetting increases in metabolic requirements. Second, even though our TEE estimates for adults in India fell by over 100 calories per day between 1983-2012, more than enough to account for the measured decline in caloric intake and equivalent to between a third and a half of the average rural-urban gap around 1998, there was substantial demographic change in the population that more than offset this effect. The share of children under age 15 fell from 40% to 30% during this period, and our analysis shows that they have both lower metabolic requirements and substantially lower activity levels, particularly in rural areas. Because the average TEE of adults is much higher than children, population-level demographic change of the magnitude that occurred in India during this period exerts very powerful upward pressure on population TEE, enough
to offset reasonably large reductions in TEE for a given age/sex. Thus, although we find
some evidence that variation in TEE can help account for the caloric-intake household size
relationship explored in Deaton and Paxson (1998), and can explain some of the national
and cross-state patterns of caloric intake in the 2005-2012 period, we conclude that it is
unlikely to generate a large reduction in TEE that could “explain” the decline in caloric
intake measured in the NSS. To the extent that there was a real decline in population
caloric intake during the 1983-2005 period, an improving disease environment (as shown
by Duh and Spears (2017)) is a more likely explanation. Our estimates of TEE are also
consistent with a large degree of measurement error in NSS caloric intake, which recent
data suggests is large enough to plausibly explain much of the apparent decline.

Although TEE is not usually considered by economists, our analysis suggests that
there are interesting insights from quantifying it. Better measures of TEE could be used
for the formulation and revision of poverty lines and equivalence scales, although given
the difficulty of accurately measuring caloric intake or TEE it seems clear that anthropo-
metric outcomes (e.g. stunting, wasting, BMI and weight gain over time) are more
reliable for the evaluation of adequate nutritional status. Estimates of TEE could also
be useful inputs into quantitative models of structural and demographic change since
“subsistence” food requirements may vary with mechanization and household composi-
tion. In addition, we have shown that quantifying TEE is a useful contribution to debates
about estimated caloric intake and nutritional outcomes – given the scope for measure-
ment error, measuring both calories in, calories out, and net caloric surplus can shed light
on biases in measurement. Time-use and anthropometric surveys are available for many
developing countries, and the potential availability of multiple waves of time-use data
in developing countries could be particularly useful for this type of exercise, as it may
allow for better quantification of changes in physical activity levels related to travel and
the intensive (per hour) and extensive (number of hours) physical margin of domestic
and market work that are harder to quantify with our methodology. Another promising
alternative to costly household surveys is the use of technology for the purpose of capturing TEE. Most modern smartphones can track physical activity levels reasonably well and simple devices that monitor heart rates offer the potential to estimate variation in activity levels for larger samples, greatly increasing the scope for estimating activity levels for populations outside the laboratory relative to the prohibitively expensive chemical methods that prevail in the nutrition literature.

References


Fiedler, John L. and Suryakant Yadav, “How can we better capture food away from Home? Lessons from India linking person-level meal and household-level food data,” *Food Policy*, 2017, 72, 81–93.


Krishnaveni, GK, SR Veena, R Kuriyan, RP Kishore, AK Wills, M Nalinakshi, S Kehoe, CH Fall, and AV Kurpad, “Relationship between physical activity measured using ac-


Figure 1: Household TEE varies with education, expenditure and household size in the cross-section

Unconditional relationship between household Total Energy Expenditure (TEE) (daily per person) and mean years of education for members over 18, log per capita expenditure (dropping observations in the 1% left and right tails), and household size (up to size 10) in the 1998-1999 Time Use Survey (TUS) cross-section. Activity Levels (AL) is measured directly and Resting Energy Expenditure (REE) is imputed using the NFHS (see text for description). Each observation represents a household average. Note that 92% of households are between 2 and 8 members.
Figure 2: Changes in caloric intake per capita and TEE by state, 1983-2012

State-level average TEE is imputed using NSS Schedule 10 as described in the text and caloric intake per capita is the authors’ estimate using NSS Schedule 1. The scatter plot excludes states with population below 10 million for readability.
Table 1: Daily Resting Energy Expenditure (REE) and Activity Level (AL) estimates

Panel A: REE (calories/day) from National Family Health Survey (NFHS)

<table>
<thead>
<tr>
<th></th>
<th>Ever marr. women 15-49</th>
<th>Women 15-49</th>
<th>Men 15-54</th>
<th>Children 0-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998 (NFHS-2)</td>
<td>1128</td>
<td>.</td>
<td>.</td>
<td>502</td>
</tr>
<tr>
<td>1998 Rural</td>
<td>1114</td>
<td>.</td>
<td>.</td>
<td>499</td>
</tr>
<tr>
<td>1998 Urban</td>
<td>1169</td>
<td>.</td>
<td>.</td>
<td>513</td>
</tr>
<tr>
<td>2005 (NFHS-3)</td>
<td>1145</td>
<td>1140</td>
<td>1496</td>
<td>492</td>
</tr>
<tr>
<td>2005 Rural</td>
<td>1126</td>
<td>1122</td>
<td>1471</td>
<td>485</td>
</tr>
<tr>
<td>2005 Urban</td>
<td>1189</td>
<td>1177</td>
<td>1542</td>
<td>517</td>
</tr>
<tr>
<td>2015 (NFHS-4)</td>
<td>1180</td>
<td>1169</td>
<td>1531</td>
<td>514</td>
</tr>
<tr>
<td>2015 Rural</td>
<td>1161</td>
<td>1151</td>
<td>1510</td>
<td>507</td>
</tr>
<tr>
<td>2015 Urban</td>
<td>1221</td>
<td>1205</td>
<td>1570</td>
<td>532</td>
</tr>
</tbody>
</table>

Panel B: AL (rest=1) from Time Use Survey (TUS) 1998-1999

<table>
<thead>
<tr>
<th></th>
<th>Male 15+</th>
<th>Female 15+</th>
<th>Male 6-14</th>
<th>Female 6-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-99 mean</td>
<td>1.93</td>
<td>1.89</td>
<td>1.49</td>
<td>1.51</td>
</tr>
<tr>
<td>1998-1999 SD</td>
<td>0.47</td>
<td>0.34</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>1998-99 Rural mean</td>
<td>2.03</td>
<td>1.96</td>
<td>1.50</td>
<td>1.53</td>
</tr>
<tr>
<td>1998-99 Urban mean</td>
<td>1.69</td>
<td>1.71</td>
<td>1.47</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Panel A is based on applying the equations in Henry (2005) to heights and weights reported in NFHS. Panel B is based on using the matched activity factors from FAO (2001) (see Online Appendix Table 1) and activities reported in the TUS. Reported values are population statistics using survey sampling weights.
Table 2: Household TEE and household characteristics

<table>
<thead>
<tr>
<th>Dep.var.(logs)</th>
<th>TEE</th>
<th>REE</th>
<th>AL</th>
<th>TEE</th>
<th>REE</th>
<th>AL</th>
<th>Cal</th>
<th>Cal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sizes</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>2-8</td>
<td>2-8</td>
<td>2-8</td>
<td>All</td>
<td>2-8</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td></td>
</tr>
<tr>
<td>Mean school</td>
<td>-0.004***</td>
<td>0.006***</td>
<td>-0.009***</td>
<td>-0.004***</td>
<td>0.006***</td>
<td>-0.009***</td>
<td>-0.008***</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Log exp. p.c.</td>
<td>-0.010*</td>
<td>0.002</td>
<td>-0.011**</td>
<td>-0.013**</td>
<td>0.002**</td>
<td>-0.014***</td>
<td>0.440***</td>
<td>0.432***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.014)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Log hh size</td>
<td>0.003</td>
<td>0.014***</td>
<td>-0.012**</td>
<td>-0.017**</td>
<td>0.016***</td>
<td>-0.033***</td>
<td>-0.015**</td>
<td>-0.029***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Observations</td>
<td>18,528</td>
<td>18,528</td>
<td>18,528</td>
<td>17,013</td>
<td>17,013</td>
<td>17,013</td>
<td>28,248</td>
<td>24,908</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.668</td>
<td>0.922</td>
<td>0.502</td>
<td>0.692</td>
<td>0.940</td>
<td>0.528</td>
<td>0.394</td>
<td>0.372</td>
</tr>
</tbody>
</table>

Standard errors in parentheses clustered by state/sector. *** p<0.01, ** p<0.05, * p<0.1. Controls include demographic ratios (male/female aged 0-2,3-5,6-9,10-14,15-17,18-59,60+), SC/ST status, religion, and sector by state dummies. Columns 1-6 use the TUS sample (1998-1999) with activity levels and estimated REE. Columns 7 and 8 use caloric intake per capita estimated for the 50th NSS round (1993-94) in the six TUS states.
Table 3: Estimates of daily TEE and its components over time

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
<th>Males 15+</th>
<th>Females 15+</th>
<th>Under age 15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TEE (1) REE (2) AL (3)</td>
<td>TEE (4) AL (5)</td>
<td>TEE (6) AL (7)</td>
<td>TEE (8) AL (9) Pop. share (10)</td>
</tr>
<tr>
<td>1983</td>
<td>2136 1179 1.77</td>
<td>2864 1.95</td>
<td>2244 1.93</td>
<td>1499 1.52 0.40</td>
</tr>
<tr>
<td>1988</td>
<td>2134 1186 1.77</td>
<td>2858 1.94</td>
<td>2234 1.92</td>
<td>1475 1.50 0.39</td>
</tr>
<tr>
<td>1994</td>
<td>2158 1200 1.77</td>
<td>2855 1.93</td>
<td>2241 1.91</td>
<td>1463 1.49 0.36</td>
</tr>
<tr>
<td>2000</td>
<td>2155 1210 1.75</td>
<td>2829 1.91</td>
<td>2225 1.89</td>
<td>1477 1.48 0.36</td>
</tr>
<tr>
<td>2005</td>
<td>2172 1222 1.75</td>
<td>2825 1.90</td>
<td>2215 1.88</td>
<td>1476 1.47 0.34</td>
</tr>
<tr>
<td>2009</td>
<td>2159 1243 1.72</td>
<td>2753 1.85</td>
<td>2182 1.83</td>
<td>1460 1.44 0.31</td>
</tr>
<tr>
<td>2012</td>
<td>2153 1250 1.70</td>
<td>2729 1.83</td>
<td>2158 1.81</td>
<td>1469 1.44 0.30</td>
</tr>
<tr>
<td>Rural sector mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>2168 1165 1.81</td>
<td>2959 2.04</td>
<td>2290 1.99</td>
<td>1505 1.53 0.41</td>
</tr>
<tr>
<td>1994</td>
<td>2190 1183 1.81</td>
<td>2951 2.02</td>
<td>2289 1.97</td>
<td>1458 1.49 0.37</td>
</tr>
<tr>
<td>2000</td>
<td>2184 1191 1.79</td>
<td>2930 2.00</td>
<td>2275 1.96</td>
<td>1472 1.48 0.37</td>
</tr>
<tr>
<td>2005</td>
<td>2199 1201 1.79</td>
<td>2930 2.00</td>
<td>2264 1.94</td>
<td>1469 1.47 0.35</td>
</tr>
<tr>
<td>2012</td>
<td>2179 1230 1.74</td>
<td>2828 1.92</td>
<td>2197 1.86</td>
<td>1455 1.43 0.31</td>
</tr>
<tr>
<td>TUS state mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>2152 1183 1.78</td>
<td>2864 1.95</td>
<td>2280 1.95</td>
<td>1490 1.51 0.39</td>
</tr>
<tr>
<td>1994</td>
<td>2184 1204 1.78</td>
<td>2857 1.94</td>
<td>2281 1.94</td>
<td>1454 1.48 0.35</td>
</tr>
<tr>
<td>2000</td>
<td>2190 1214 1.77</td>
<td>2847 1.93</td>
<td>2264 1.92</td>
<td>1462 1.47 0.34</td>
</tr>
<tr>
<td>2005</td>
<td>2208 1224 1.77</td>
<td>2847 1.93</td>
<td>2260 1.91</td>
<td>1468 1.47 0.32</td>
</tr>
<tr>
<td>2012</td>
<td>2199 1254 1.73</td>
<td>2776 1.87</td>
<td>2196 1.83</td>
<td>1452 1.42 0.28</td>
</tr>
</tbody>
</table>

TEE is Total Energy Expenditure, REE is Resting Energy Expenditure, and AL is activity level. See text for a description of how we estimate these objects using the NFHS, TUS, and NSS schedule 10 (employment survey). Columns 1-3 report population means (using sampling weights). Columns 4-9 are population means for subsets of the population. Column 10 is the share of the children aged 14 and under in the population. The reported years correspond to the last year covered by NSS “thick” survey rounds (38, 43, 50, 55, 61, 66, 68).
Table 4: Comparison of daily TEE, caloric intake, caloric burden of disease, and caloric surplus implied by weight gain

<table>
<thead>
<tr>
<th></th>
<th>Consumption</th>
<th>TEE</th>
<th>Disease</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real mpce</td>
<td>NSS cal</td>
<td>NSS cal 7-day</td>
<td>Share 7-day</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>1983</td>
<td>302</td>
<td>2190</td>
<td>2307*</td>
<td>0.37</td>
</tr>
<tr>
<td>1988</td>
<td>305</td>
<td>2202</td>
<td>2337*</td>
<td>0.42</td>
</tr>
<tr>
<td>1994</td>
<td>321</td>
<td>2133</td>
<td>2269*</td>
<td>0.43</td>
</tr>
<tr>
<td>2000</td>
<td>350</td>
<td>2151</td>
<td>2290*</td>
<td>0.44</td>
</tr>
<tr>
<td>2005</td>
<td>355</td>
<td>2040</td>
<td>2184*</td>
<td>0.48</td>
</tr>
<tr>
<td>2010</td>
<td>386</td>
<td>2000</td>
<td>2140</td>
<td>0.48</td>
</tr>
<tr>
<td>2012</td>
<td>479</td>
<td>2087</td>
<td>2225</td>
<td>0.49</td>
</tr>
</tbody>
</table>


Column 1 is our calculation of real per capita expenditure using NSS schedule 1. Columns 2 is the official estimate of caloric intake from NSSO (2013) using the 30-day recall period (we use Deaton and Drèze (2009) for 1988 as this is missing in the report). Column 3 is the official estimate using the new schedule type 2, which uses a 7-day recall for all food except cereals, dairy, and pulses. * indicates our imputation, based on the share of expenditures on the 7-day recall items (column 4) and the difference in caloric intake between the 30-day and 7-day schedule. Column 5 is our independent estimate of caloric intake (see appendix). Columns 6 and 7 are our population and adult TEE estimates. Columns 8 through 10 are based on the reported coefficients for the calorie-disease slope in Duh and Spears (2017), where disease is proxied by infant mortality. Column 9 reports infant mortality rate (IMR) for India (per 1,000 live births) from the World Development Indicators and the mid-level estimate based on a slope coefficient of one. Columns 8 and 10 multiply IMR by 1.74 and 0.426, the high and low end estimates reported in Duh and Spears (2017) Table 8. Column 11 “Surplus” is the net caloric surplus consistent with average weight gain using the NFHS data (see text and Online Appendix Table 8).
Data Construction Appendix

Resting Energy Expenditures (REE)

The best anthropometric predictor of REE is fat-free mass, but as this is rarely measured in the field, prediction equations have been developed based on regressing laboratory measured REE on more commonly observed variables (St. Jeor and Stumbo (1999)). A standard approach, adopted by FAO (2001) and Indian Council of Medical Research (2009), is to use prediction equations that regress REE on gender, age, weight, and in some cases height. Estimating separate equations by age and sex, and incorporating height, helps account for systematic average differences in fat-free mass, e.g. women, older, and shorter individuals tend to have lower fat-free mass for a given weight. We use the equations provided in Henry (2005) that include height and weight for several age/sex groups. However, our results are not very sensitive to the use of alternative prediction equations. In Online Appendix Table 2 Panel A, we generate REE estimates using the prediction equations used by Indian Council of Medical Research (2009). We have also obtained similar results using the prediction equations from St. Jeor and Stumbo (1999). There have been laboratory measurements of REE in India but none for a large or representative sample of individuals and none generating specific predictive equations. Ferro-Luzzi et al. (1997) attempt a validation of the FAO equations on Indian data and find a reasonable fit.

Our preferred method for imputing REE to individuals in other data sets is to regress the REE predicted by the Henry (2005) equations using the age, sex, height and weight from the NFHS on a set of common variables available in the NSS Schedule 10 employment survey. Because age and sex are always observed, this procedure generates additional variation in TEE conditional on age and sex that captures average differences in height and weight associated with socioeconomic status, maternal and childhood nutrition. Online Appendix Table 3 reports the variables and coefficients we use for these regressions, which include individual and household age, education, demographics, school attendance, land holdings, and two-digit primary occupation code (using NCO1968 clas-
sification and available concordances for the later NFHS rounds). We estimate separate equations by sector and sex, pooling the available years of data (which includes 1998 for women and 2015 for both men and women) and including year fixed effects (the omitted category is 2005). In principal we could also use separate predictive equations for height and weight, and then apply the Henry (2005) equation, but we find the results are similar.

Note that we estimate these equations only for the adult age ranges in the NFHS eligible for height and weight measurement (15-49 for women and 15-54 for men). For individuals outside of this range, we adopt the following procedure:

1. We assume that their age is equal to 15 or 49 (54 for men) and use the predictive equation for REE

2. We scale this REE by the ratio of average REE for that individual’s age and sex relative to the average REE for their “assumed age” used in the predictive equation.

To scale REE, we use the age and gender specific mean height and weight reported by the National Nutritional Monitoring Bureau for rural areas of 16 Indian states between 2000-2002 (reported in Indian Council of Medical Research (2009)), together with the Henry (2005) equations. Note that Indian Council of Medical Research (2009) uses the 95th percentile of this same distribution to generate their recommended daily caloric intake, highlighting an important difference between their prescriptive estimates of recommended daily caloric intake and our descriptive estimates of population total energy expenditure.

We make two additional adjustments in our calculations. First, because children under 18 are still growing, we add the calories consistent with “normal” growth, i.e. the pattern of observed weight gain. We use the 2 kilocalories per gram of tissue synthesis suggested in FAO (2001) combined with the average annual weight gain for each age and gender in the National Nutrition Monitoring Bureau data. We are unable to make adjustments for the additional energy required by pregnant or nursing mothers due to lack of data, but we do include the energy requirements of infants themselves. We set REE for infants
(children under age 1) such that male babies have TEE of 650 calories and female babies have TEE of 600 calories, consistent with the Indian Council of Medical Research (2009) report. Second, we replace values outside of the 1% of the tails of the distribution with the values at these percentile cutoffs prior to estimation.

**Activity Levels (AL)**

Online Appendix Table 1 lists the time-use survey activity codes and headings, the average share of time spent on the activity by rural and urban individuals, and the matched activity factors from FAO (2001). Our results are not very sensitive to using alternative sources of activity factors, such as those developed for richer countries that tend to contain more detailed classifications of sports and exercise activities. Online Appendix Table 1 includes the matched activities and activity factors from Ainsworth et al. (2000) (used by Cutler et al. (2003)) for comparison, and Online Appendix Table 2 Panel C shows that the difference in activity levels using our preferred measure and using this alternative set of activity factors is quite small.

We make two additional assumptions about activity factors to derive activity levels. First, since the time-use data reports time spent on “related” or “other” activities within certain headings, we need to assign an activity factor in these cases. We assume that the activity factor in this case is equal to the average for the activities we do match under the same heading. Second, since we do not observe mode of travel, we assume a single activity factor equal to 3 for all travel. We view this as reasonable given that car ownership is very low in India and many other forms of transport (including bicycles, motorcycles, animal-carts and public transit) are equally and sometimes more energy intensive than walking. For example, FAO (2001) reports activity factors for “sitting on a bus/train” (1.2), “driving a car/truck (2.0), “walking around/strolling” (2.1), and more intense activities like “carrying a 20-30kg load on head” (3.5), “walking quickly” (3.8), and “cycling” (5.6). Our assumed activity factor for travel of 3 lies between these extremes, and it is in the same range as “walking slowly” and “driving a motorcycle.” Based on NSS data, the
car ownership rate in rural areas as late as 2005 was only 1%, and in urban areas was only 3%. Motorcycle and bicycle ownership rates are higher in 2005 (8% and 48% respectively in rural areas, 25% and 44% in urban areas). Our approach allows better-off households (proxied by per capita expenditure) to economize on travel time by using faster modes of transport, which leads to lower activity levels since travel is fairly energy-intensive compared to most other activities, but does not allow us to capture the potentially higher or lower activity factors associated with different modes of transport for a given amount of travel time measured in the TUS.

To impute activity levels to individuals in the NSS-E, we regress the individual activity levels we calculate using the TUS on a set of common variables. We estimate separate regressions for each combination of rural/urban, male/female, under 15 and over. Online Appendix Table 4 presents the set of variables and regression coefficients for these regressions, which include detailed age and household demographics, education levels, land, per capita expenditure, household and individual two-digit NIC (1987 National Industry Classification) and NCO (1968 National Classification of Occupations), and work status variables (e.g. self-employed, casual worker, salaried worker, student, domestic duties, etc.). We also construct dummy variables for the TUS to match the detailed agricultural task variables (based on the previous week) and domestic chores (for women with primary status as domestic), based on whether households report any time spent during the last week on activities like weeding, harvesting, caring for animals, collecting firewood or water, food collection and preparation, etc. As we did with REE, we replace values outside of the 1% of the tails of the distribution with the cutoff values prior to estimation. For individuals younger than 6, we set their activity level equal to the sample mean for 6 year olds, which is 1.46.

**Caloric intake**

For our caloric intake estimates, we use the NSS consumption survey (schedule 1) for the same years as the NSS-E. For 2010 and 2012 we use the comparable 30-day recall
period. Household respondents are asked to recall the total quantity and expenditure for each food item from a detailed list over the previous 30-days. We begin with the standard methodology of directly converting quantities of each food into calories using the calories per unit weight for each item reported in Gopalan et al. (2004), which we supplement with data from the MedIndia web-site.\textsuperscript{10} This covers 90.8\% to 95.4\% of average household food expenditure for the periods we consider.

For goods that have missing units or that are classified as “other” within a major category heading (e.g. fruits, vegetables, meat, dairy), we assign calories proportional to the average calorie per rupee calculated within that category and NSS region. This increases the share of food expenditures covered by an additional 0.5\%. The rest of food expenditures fall under the processed food and beverages categories, where the lack of units and/or vague classification make the use of a caloric conversion difficult if not impossible, e.g. the count of cooked meals or “cold beverages bottled/canned” or unitless expenditures on “prepared sweets” or “salted refreshments.” Calories from reported tea, coffee, and alcohol consumption are straightforward to incorporate because their units are reported. For the other goods, we assume that calories per rupee are equal to 50\% of the calories per rupee that can be directly converted across all goods. The data we could find indicates ingredient costs make up 40\% of the sale price at large Indian restaurants (Federation of Hotel and Restaurant Associations of India (2004)) and the value for richer countries is typically in the 20-35\% range. A value of 50\% puts the calories/rupee of processed foods, beverages and cooked meals roughly equal to the dairy category, while a value of 66\% puts it equal to pulses or pure sugar.

The other adjustment we make is similar to the official NSS estimates, which incorporate survey information on the number of meals consumed by household members at home and away on payment (which are drawn from the calories measured in the survey), the number of free meals consumed away from home through employers, schools

\textsuperscript{10}See http://www.medindia.net/calories-in-indian-food/index.asp.
and other households (which are not included in the survey quantities), and meals given to non-household members (which are drawn from the survey quantities but do not contribute to household caloric intake). We use the simple adjustment factor based on the formula \( \text{adjustment factor} = \frac{\text{meals at home} + \text{meals away from home free}}{\text{meals at home} + \text{meals to others}} \). We therefore assume that households that consume more calories per meal at home both give and receive free meals that are proportionately higher in calories, and that free meals given and received enter symmetrically. Our reported estimates censor the estimated distribution at 960 (equivalent to the 1% cutoff in 1983) and 4400 (the 99% cutoff in 2012), which lowers the estimated level of intake for some years due to the presence of large right tail outliers but has a minimal effect on the trends.