Sounding a Note of Caution  
Data on MGNREGA in Tamil Nadu

GRACE CARSWELL, ROB CRIPPS

Official data from the Mahatma Gandhi National Rural Employment Guarantee Act website is examined alongside independently collected survey and ethnographic data. The validity of the official data is verified by comparing it to field-level observations. This is done at various levels, and in doing so, it is shown that in certain ways the official MGNREGA data is robust. In other very important ways the data is shown to be highly problematic for these particular villages in Tamil Nadu, for which the exercise is done.

India’s flagship Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) is an unprecedented nationwide employment scheme that already benefits millions of rural households across the subcontinent. The National Rural Employment Guarantee Act (NREGA) was enacted by Parliament in 2005, rolled out across a selection of districts in 2006 and extended across all districts in 2008. The scheme seeks to provide basic social security to India’s rural poor and provides 100 days of guaranteed waged employment per year to every rural household. Since it was rolled out across India, MGNREGA has been marked by a huge diversity in implementation and impact, and there is a growing body of literature on the scheme. Thus, for example, there has been research on patterns of uneven implementation (Dreze and Oldiges 2011; Ramachandran and Rawal 2010), on processes of monitoring and auditing (Menon 2008; Aiyar 2010), on the effects of NREGA on wages (Berg et al 2012; Imbert and Papp 2012) and on the scheme’s poverty alleviation potential (Dreze and Khera 2009).

Much of this research has used official MGNREGA data (available on www.nrega.nic.in) to look at, for example, regional patterns of uptake (Dreze and Oldiges 2011) or the performance of the scheme over time (Jha and Gaiha 2012). A number of studies have used the official MGNREGA data alongside other data sets to evaluate different aspects of the scheme. For example, Dutta et al (2012) use it alongside National Sample Survey data to explore how it is meeting demands for work, while Dasgupta and Sudarshan (2011) similarly use MGNREGA data alongside NSS data to study gender wage gaps. Jeyaranjan (2011) draws on official data of a number of households registered for MGNREGA work, and their sex, age and caste composition, as well as number of days of employment in order to explore the effects of the policy on women in Tamil Nadu. These studies have taken the MGNREGA data as correct and reliable at face value without validating their reliability by triangulation with other types of data.

Other literature has used surveys or other field-based research undertaken by the researchers to explore MGNREGA and its impacts. Thus, for example, Dreze and Khera (2009) surveyed MGNREGA workers in order to review the first years of the scheme, while Khera and Nayak (2009) draw on the same survey to focus on the effects of MGNREGA on women. A number of studies have conducted their own surveys to explore women’s participation in MGNREGA and the empowerment effects of the scheme on women (see for example Pankaj and Tankha 2010; Sudarshan et al 2010; Carswell and De Neve 2012b). Other studies have similarly used their own surveys to look more broadly at the success of MGNREGA as a social protection policy (see Holmes et al 2010 for Madhya Pradesh; Carswell and De Neve 2012a for Tamil Nadu).

Few articles, however, have attempted to bring together the official data with surveys and other field-based data collected by the researchers themselves, and in doing so assess the quality of the official data. In our article we will do just this: comparing field-level observations and surveys with official data. We will do this in different ways. First, we will compare the data in terms of absolute numbers of people registered as having attended for work, drawing in particular on the muster rolls available on the MGNREGA website. Second, we will look at the individual characteristics of MGNREGA workers, comparing information available on the MGNREGA website with data collected in the field from the same workers.

Setting the Scene

This article brings together two very different types of data. First, it draws on data from the official MGNREGA website. For our purposes the most important parts of this data set are the muster rolls and the job card data. Muster rolls are lists of who worked on a particular job...
on a particular day. The muster roll identifies people by a job card number, and this then links to the job card data, which includes further details about individual workers. Second, the article draws on ethnographic fieldwork carried out in two villages located in two different blocks in rural Tamil Nadu in November-December 2011. Both these villages, which we call Allapuram and Mannapalayam, were the sites of previous fieldwork (conducted in 2008-09) during which we had built up a broader picture of livelihoods, employment opportunities and social relations in the region. The fieldwork conducted in 2011 was designed to gather quantitative and qualitative information about MGNREGA and its effects on labour markets and livelihoods. A short survey was conducted with 109 MGNREGA workers (55 from Allapuram and 54 from Mannapalayam) which included information on occupations, incomes, caste, education, debt and sources of credit. Spending a month with workers at MGNREGA worksites as they cleared roadsides and irrigation canals we used a combination of participant observation, case studies, focus group discussions and in-depth interviews with workers about the scheme and its impacts on their lives. In addition we carried out interviews with site organisers, village administrators and block-level officials.

The villages of Allapuram and Mannapalayam are located within 20 kilometres of the booming garment town of Tiruppur. Allapuram has a thriving agricultural sector, and also sends commuters, from across castes, to work in Tiruppur. Mannapalayam is predominantly a power loom village, and has attracted many migrants to work in the powerloom sector (Carswell 2012). In both villages MGNREGA has been implemented “successfully” by several standards, and the scheme particularly benefits women and dalits (Carswell and De Neve 2012a; Carswell and De Neve 2012b).

Assessing the Data

Registries and Muster Rolls: During our fieldwork we copied out the registries (filled in by the local in charge or supervisor) for each of the village worksites for two weeks. The registries included the numbers of people working each day, their gender and their caste (scheduled cast (sc) or “Other”). We corroborated the registries by doing a headcount of workers attending, and were confident that for these weeks the numbers in the handwritten registries were true and accurate. How, then, do these registries compare with the data online? Online information about daily attendance is found in the muster rolls which are broken down for different jobs. A job would be, for example, clearing the roadsides between two named villages, and would normally continue over several days or even weeks. A muster roll lists everyone who worked on that job during the week (from a Friday to the following Thursday), and a new muster roll is begun if the job continues into the following week. These muster rolls act as the primary data source from which reports are generated on the MGNREGA site and evaluating the muster rolls against data collected in the field is a key part of verifying the official MGNREGA data.

We compared the handwritten registries of attendance (collected in the field) with the online recorded daily attendance (collected from the online muster rolls), and Table 1 shows the two sets of data for each of the villages, showing both daily and weekly totals.

Table 1 shows the comparison between the official and field data, indicating high levels of accuracy in the data from Mannapalayam but notable errors in the data from Allapuram. Whilst there is evidence of some minor discrepancies in the daily figures recorded on the muster rolls from Mannapalayam, the majority of days were accurate to within 3% of the observed numbers working. The table also shows that the weekly totals were completely accurate. However, in Allapuram there were considerable inaccuracies in the weekly totals shown on the muster rolls, which were even more evident in the daily figures recorded. Allapuram’s muster rolls for those two weeks show the same individuals (55 for the first sample week and 53 for the second) working every single day. However, we know from our observations that the same individuals were not working every day of the week. Indeed, the field data collected for those two weeks reveals variations in the numbers working each day (from a low of 38 up to a peak of 63). This means that not only are the...
weekly totals incorrect, but all reports drawn from the data for that village (for example, reports showing employment provided over time) also contain the same errors.

Job Cards and Field Survey: The muster rolls list all the individuals working on a particular job, recording their name and job card number, and providing a link to their online job card, which contains information on age, gender and caste. The existence of such data on the job cards, and the fact that muster rolls are linked to job cards mean that, potentially, this data set could be used to examine the characteristics of MGNREGA workers, an exciting prospect for researchers. We had hoped to be able to explore, for example, whether men or women worked more during particular seasons, linked perhaps to the seasonality of agriculture, or whether the characteristics of MGNREGA workers changed at particular moments in time, such as during industrial action in the rural power loom industry.

However, before doing this, and to assess the quality of the data, we examined the characteristics of workers, comparing the online job card data with information we had collected in the field. Because we had found that the recording of workers’ attendance on the muster rolls for Allapuram was inaccurate, we decided to look only at Mannapalayam for this stage of the analysis. Working through the muster rolls for each of the two sample weeks in Mannapalayam, the age, caste and gender of each worker was recorded alongside the days they worked. In Table 2, this data is compared with the workers’ caste and gender characteristics as collected from the field registries.

Table 2 reveals that sizeable errors exist across the sample weeks with regards to both gender and caste, with the online data under-representing the number of women and SCs in each of the weeks.

This under-representation appears to stem from different errors. First, errors in the recording of who has actually worked from within each household on the online system, and secondly errors on the job cards themselves. In order to understand these errors we first need to explain the difference between the physical job card and the online job card (and how each of these are numbered). Every worker holds their own physical job card. It is numbered with a household number, followed by a letter (A, B, C, etc) with each letter representing an individual member of that household. Online, however, a single job card, ending usually in an A, includes the names of all the individuals within a household who have registered for a job card and each individual within the household is not allocated their own letters. In our survey of 109 workers we have both the physical job card number and the online job card number for 98 of them. Of the physical job cards there were 59 ending in A, 30 ending in B, six ending in C and three ending in D. But online almost all (96 out of 98) ended in A, with just two ending in B. There were no C5 or D5 listed at all. It is clear therefore that while the lettering of individual job cards is practised in the field, it is not carried through to the online system. While muster rolls list the names of workers, the online job card number can only be used to identify which household has worked, but not which individual has worked.

A further error appears relating to who worked, that is to whether the correctly named individuals were identified as working on any particular day. Of the 109 individuals who we surveyed across the two villages, we have the physical job card details of 103. Of these 103 individuals, 60 were correctly listed online as working on the days we surveyed them. Of the remainder, 26 had a member of their households listed, but were not themselves listed. This is largely because of the problem with the lettering of online job cards, referred to above. However, in addition to this error, we found that a further 17 individuals were not on the muster roll as having worked that day at all – neither them, or anyone else in their household. All but one of these 17 errors was from Allapuram, where we know the muster rolls were highly unreliable.

Errors were also found in relation to the individual characteristics of workers recorded on the online job cards. As part of our survey, we collected basic demographic information of workers, and are able to compare this data with that on the online job cards.

Table 3 shows that there were significant discrepancies between the online data and that collected in the field surveys. In relation to gender the error is not too bad, but this error appears in addition the error discussed above of the incorrect lettering of job cards, and the incorrect recording of which individuals within a household are working. Combined these errors mean that we cannot be confident about data on gender in the online data set. In relation to caste, it is clear that there are many errors – 21 out of 100 workers had their caste information incorrectly recorded. It is clear therefore that the caste information in this data set is unreliable. In relation to age we again found very significant errors. While we can expect estimations of age to be inaccurate, the errors that we identified were too large to simply be poor estimations. Of the 100 individuals 77 had an online age of within 10 years of their surveyed age, meaning there
were 23 who were more than 10 years out.6 Of these remaining 23, 15 had an online age of more than 15 years out, and 11 of them had an online age of more than 20 years out. To give just one example: Sivagami who had told us she was a 66-year-old widow (and her life story confirmed this) had an online job card entry saying she was 34 years old. Most of the errors in the online data underestimated people’s ages – indeed only one of the people whose age was 10 years or more out was older than in reality, with the remainder all being given younger ages on the online data.

Conclusions

This article has looked at different aspects of the official MGNREGA data available online, and which is widely used by researchers. It has compared it to field observations and survey data collected by one of the authors. It shows that parts of the official online data are fairly robust. For one of the two villages the attendance data online was very close to that observed in the field. For the other village, however, it is clear that the online attendance data is inaccurate: the online data set suggests that the same individuals worked every day of the week, when we know this was not the case. This particular inaccuracy is extremely easy to spot, and researchers would do well to carry out checks such as this one before using the attendance data.

Our observations suggest that in the field the correct worker (with job card number and letter) was assiduously recorded, whereas online job card numbers refer simply to the household. There is an obvious reason for this: at a local level it is critically important that the correct individuals are identified as working so that the weekly payment can be made to the correct individual within a household. On the online system, however, no attention has been paid to identifying which individuals within a household are working, as from an administrative perspective this is unimportant. At this level it is important for the administration to keep a track of how many days each household has worked (as the limits on MGNREGA work of 100 days per year apply to the household and not the individual). The correct recording as to which member of the household undertook the work, on the online system is of no material consequence to the monitoring of the 100 days limit or to the budgets spent and therefore errors can be more easily tolerated.

But the real inaccuracies in the data become apparent when we look at the data on individual characteristics of MGNREGA workers. The use of letters and numbers to identify individuals on the ground is not carried through to the online system. This, combined with the fact that there are many instances of incorrect household members being listed as working, means that it cannot be reliably inferred from the online data set as to which individual within a household works. Finally, the actual information on the job cards is also unreliable: age and caste details have particularly high levels of inaccuracy. These findings lead us to suggest that there are simply too many errors in the online data on individuals’ characteristics for this to be usable by researchers. We do not suggest that this is a deliberate policy to skew the data in any particular way. But we do wish to sound a note of caution to other researchers intending to use this data, and in particular the individual characteristics part of the data set.

NOTES

1 These, and all other names, are anonymised.
2 Both periods of fieldwork (2008-09 and 2011) were undertaken by Carswell and De Neve. In the 2008-09 research we had carried out a year of in-depth fieldwork, collecting detailed ethnographic material as well as survey data on all 240 households in Allapuram and 279 households in Mannapalayam.
3 Both villages had about the same level of error with 12 of these individuals being from Allapuram and 14 from Mannapalayam.
4 If we narrow the range and look at those with an online age of within five years of their surveyed age we find that 57 of the 100 individuals for whom we had information have an online age of within five years of their surveyed age (meaning that 43 were more than five years out).
5 Error of plus or minus more than 10 years.
6 If we narrow the range and look at those with an online age of within five years of their surveyed age we find that 57 out of the 100 individuals for whom we had information have an online age of within five years of their surveyed age (meaning that 43 were more than five years out).

REFERENCES