A comparison of four approaches for measuring clinician time use

JOHN H BRATT,1 JAMES FOREIT,2 PAI-LIEN CHEN,1 CAROLINE WEST,1 BARBARA JANOWITZ1 AND TERESA DE VARGAS3

1Family Health International, Research Triangle Park, North Carolina, USA, 2FRONTIERS Project, The Population Council, New York, and 3Centros Médicos de Orientación y Planificación Familiar (CEMOPLAF), Ecuador

Concern about rapid growth in demand for reproductive health services in developing countries has created interest in productivity and costs of existing programmes. Staff costs usually constitute the largest share of total service costs, meriting special effort to ensure that they are measured accurately. Several techniques have been used in the literature to analyze staff activity, but these techniques have not been validated. This paper reports on a study conducted in three Ecuadoran clinics. The study uses an observational time-motion (TM) technique as a benchmark, and compares results from three other techniques to those obtained using TM. None of the alternative techniques produces estimates that agreed with TM estimates; deviations from TM are particularly large for non-contact time, defined as clinician activities carried out when clients are not present. Implications of these findings for productivity and cost studies are discussed, and possible avenues for future research are proposed.

Introduction

In recent years, donors, governments and others in the international health community have focused increased attention on the costs of providing reproductive health (RH) services in developing countries. As ever-larger cohorts of women enter their reproductive years, demand for services is growing at a faster pace than resources needed by programmes to produce services. Several studies have made projections of the global cost of meeting future demand for RH services, and although their estimates vary, all agree that massive increases in spending will be needed (Janowitz 1993). In response to the uncertainty of future funding, some programmes are experimenting with ways to generate local income and to make better use of existing resources.

Productivity of current programme resources is an important determinant of future costs of RH services. If clinic resources such as labour and capital are nearly fully utilized at current levels of output, clinics will need additional labour and capital to meet higher demand, and costs of services may increase substantially. But if labour and capital are under-utilized, additional clients can be served without added investment in labour and capital. Evidence from some RH programmes in the developing world suggests that programme personnel are under-utilized; for example, two recent studies in Cameroon and Bangladesh revealed substantial unallocated time among rural fieldworkers (Bryant and Essomba 1995; Janowitz et al. 1997), while a series of cost studies carried out in Honduras, Mexico, El Salvador and Bangladesh has shown the same situation to be true in clinics (Bratt et al. 1993, 1997; Hubacher et al. 1997; Janowitz et al. 1997).

Few RH programmes collect information on how staff divide their time between direct client care and other activities. As a consequence, programme managers are largely unaware of the amount of excess service capacity that could be used to serve additional clients. Excess capacity is estimated by analyzing provider actions that do not involve direct client contact. In the context of clinic-based service delivery, these activities can be divided into two major groups: firstly, administrative tasks and other necessary complements to client–provider interaction, such as review of client medical records, staff meetings, and preparation and clean-up of the area where clients are served; and secondly, all other staff activities that are not related to client care or clinic administration, including personal breaks, conduct of personal business, and time spent waiting for clients. Excess capacity, if any, is found within this second group of activities, and is defined as the time remaining after subtracting a reasonable amount of time for approved personal breaks.

Excess capacity can be estimated using a variety of techniques, ranging from relatively simple and straightforward methods like provider interviews or self-administered timesheets, to more complicated time-motion (TM) studies. The observational TM study is widely regarded as the ‘gold standard’ of time measurement. Developed by Frederick Taylor (1911) as a way to increase efficiency of industrial production, this technique requires direct observation of workers and careful documentation of individual actions. As this method gained popularity and spread to fields outside industrial engineering, it has evolved into a useful tool for measuring time use in a variety of settings. In the US health sector, Lurie used a TM methodology to measure time use among 35
house staff in teaching hospitals in Hennepin County, Minnesota (Lurie et al. 1989). The National Center for Health Services Research also used this technique to examine resource requirements in nursing homes (Roddy et al. 1987). TM has also been used in less developed settings; for example, Sauerborn and Reid both used the technique to study costs of maternal and child health facilities in Burkina Faso and Papua New Guinea respectively (Reid 1986; Sauerborn 1989). Although TM studies are thought to generate accurate estimates of time use, they can be costly because they require highly skilled observers that can correctly interpret staff activities. A potential disadvantage of TM is that providers know that their activities are being monitored, and may alter normal activity patterns.1

The literature on health services in developing countries also provides examples of staff time measurement using other approaches. Patient flow analysis (PFA), a technique developed by the Centers for Disease Control (CDC) to improve client flow in clinics, has been used to measure the duration and costs of provider-client interactions (Bratt et al. 1993). Other analyses of health worker time in the literature rely on provider recall, either in the form of structured interviews (Valadez et al. 1990) or self-administered timesheets (Chernichovsky and Anson 1989). One study measured length of family planning visits by asking clinic managers to estimate the duration of activities that occur during family planning visits (Stewart et al. 1997). Although provider interviews and self-administered timesheets may require less effort than observational studies, there is concern that reliance on provider self-reports may not yield valid results.

Current concerns about sustainability of RH services in developing countries ensure that accurate cost information will become increasingly important to programmes, governments and donors. Personnel costs account for a large share of total costs in most RH programmes. If methods used to measure staff time are not valid, then cost estimates based on these data will be inaccurate. The objective of this study was to compare results of a TM study with results of three alternative methods of time measurement, in order to determine the level of agreement between TM and other methods. If a less complicated method could be shown to produce similar results to TM, it would make future studies of time-use and costs easier to conduct.

Methods

The study was conducted in three clinics of the Centro Médico de Orientación y Planificación Familiar (CEMOP- PLAF), an Ecuadorian non-governmental organization (NGO) that provides reproductive health services nationwide. Study sites were selected to include clinics with diverse client characteristics and service volume. The first clinic is located in the coastal city of Esmeraldas, and has one of the largest caseloads of all CEMOPPLAF clinics. A lower-case-load clinic is located in the highland city of Otavalo, and serves a population that is largely rural and indigenous. The third clinic, known by its Spanish acronym COP (Colegio de Obstetrices de Pichincha), is located in a working-class district of Quito, the capital city; its caseload is lower than Esmeraldas, but higher than Otavalo. All three clinics offer the same services, including family planning, gynaecology, paediatrics and pre- and post-natal consultations. Providers in the clinics include physicians and nurse-midwives, most of whom work 4-hour shifts. In each facility, the study was carried out in a single consultation room and included two different providers, one working during the morning and the other in the afternoon.

Time-measurement techniques

The four approaches used to measure staff time included direct observation of clinicians, self-report of clinician activities and one hybrid technique.

TM study

A trained observer followed the clinician for the entire shift, recording actions at three-minute intervals. The observer used a digital wristwatch with an alarm that emitted a tone at the end of each interval. A pre-coded form was used to record what the clinician was doing at that exact moment. Activity categories included client contacts (e.g. IUD insertion, gynaecology check-up), work-related administrative duties, personal breaks and time spent waiting for clients.

PFA supplemented with an interview

PFA is a data collection and analysis package developed by the CDC (CDC 1993). The purpose of PFA is to document patient flow and personnel utilization in clinics. Upon entering the clinic, each client was given a pre-coded form to carry during the visit, and was instructed to present the form to all clinic staff with whom the client interacted. Clinic staff used synchronized watches to record the time at the beginning and the end of each contact; completed PFA forms therefore comprise a series of timed client–provider contacts. PFA measures total length of visits, length of individual contacts, and the total amount of time providers spend with clients, but does not provide any information on staff activities when clients are not present. To measure and classify clinician time not involving client contact, the PFA was supplemented with brief structured interviews with clinicians at the end of each shift.

Provider interviews (PI)

Study staff interviewed clinicians at the end of each shift to elicit information on time spent on specified activities. The instrument included questions about client contacts and non-contact activities. Routine programme statistics were used to help clinicians remember which types of visits they had attended that day. For each type of service provided, clinicians were asked to estimate an average duration of client contact. The remainder of the interview focused on provider activities not related to direct client care, and asked providers to estimate time spent on these activities.

Self-administered timesheets (ST)

Clinicians were provided with a logbook to record all activities carried out during the shift, regardless of whether the
activity was related to client care or not. Providers were asked to fill out the form during the shift, and to record a brief description of each activity, the time the activity began, and the time it ended.

Data collection

Figure 1 shows how the fieldwork was scheduled. The study team collected data in the three clinics over a 9-week period. In order to minimize interruptions in normal clinic routines, the study team spent one week at a time collecting data in any given clinic. We assumed a priori that the TM approach was the ‘gold standard’ for time measurement because it requires direct observation of clinician activities; therefore, TM was carried out during the entire period of data collection. Each week, one of the three other methodologies was paired with TM. The paired methodologies were implemented for 5 days (ten 4-hour shifts) in each clinic, resulting in 30 observations (3 clinics × 10 shifts) for each pair.

Classification of time spent

Each of the four techniques generated estimates of the number of minutes clinicians spent on activities during the shift. We classified minutes into three categories:

1. **Contact time** included all time spent in direct contact with clients, either individually or in a group;
2. **Non-contact productive time** included administrative activities, preparation and clean-up time related to client visits, and any other job-related tasks not involving client contact;
3. **Non-productive time** included time spent waiting for clients, breaks, absences and non-job-related activities performed during the clinic session. Some activities classified as non-productive time are unavoidable and/or necessary (i.e. coffee breaks and bathroom breaks);

Specification of outcomes

Two outcomes were considered in the analysis, both of which were derived by comparing the results obtained from TM with results from the other approaches. A continuous outcome (see below) was used to reflect the differences in provider time as measured by TM and the other three methods, while a dichotomous outcome was used as a measure of agreement between TM and the other approaches. These outcomes were evaluated for the three categories of clinician time listed above, as well as for total time spent per shift.

(a) Continuous outcome

The continuous outcome for each category of provider time was calculated by subtracting the number of minutes recorded by TM from the number of minutes recorded by the paired methodology on that shift. The absolute value of the difference was used, since our interest was the magnitude of any differences between instruments, not the direction of the difference. To test for variation in times reported by different instruments, analysis of variance for repeated measures was conducted. Our first analysis revealed that shift and clinic site were significant sources of variation. A multilogistic regression was carried out that controlled for the effects of these covariates, thereby isolating the effect of the instruments as the source of variation in the observed differences between TM and the other approaches.

(b) Dichotomous outcome

The dichotomous outcome for each category of provider time was calculated by determining whether the difference between TM and the matched approach was within 10% of the time recorded on the TM instrument. The 10% cut-off was chosen as an arbitrary limit of accuracy. If the value fell within this range, the observation was defined as consistent with TM; if not, the observation was considered to be inconsistent. A generalized linear model was used to estimate odds ratios for consistency of measurement between TM and the other three approaches.

Results

Table 1 presents findings on the average number of minutes spent per clinic shift on the three categories of provider time, as recorded by the four methods of time measurement. The number of observations for TM was three times higher than for the other methods because the study design required the TM technique to be administered during all comparisons.

The last row of the table shows that there were large disparities in the mean total time measured by the four methods. PFA and the ST recorded approximately 32% fewer minutes per shift than did TM, differences that are largely explained by underestimates of non-productive time (see second-to-last row); TM recorded an average of six to 12
Measuring clinician time use

Table 1. Mean minutes spent on activity categories per shift, by time measurement instrument

<table>
<thead>
<tr>
<th>Activity category</th>
<th>Instrument</th>
<th>TM (n = 90)</th>
<th>PFA (n = 30)</th>
<th>PI (n = 30)</th>
<th>ST (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td></td>
<td>136 (36)</td>
<td>128 (35)</td>
<td>180 (70)</td>
<td>127 (41)</td>
</tr>
<tr>
<td>Non-contact productive</td>
<td></td>
<td>38 (27)</td>
<td>20 (18)</td>
<td>33 (21)</td>
<td>17 (31)</td>
</tr>
<tr>
<td>Non-productive</td>
<td></td>
<td>50 (42)</td>
<td>4 (10)</td>
<td>6 (12)</td>
<td>8 (18)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>224 (30)</td>
<td>152 (33)</td>
<td>219 (78)</td>
<td>152 (45)</td>
</tr>
</tbody>
</table>

Values in parentheses are the standard deviation of time spent.

TM = time-motion; PFA = patient flow analysis; PI = provider interview; ST = self-administered timesheet.

times as much non-productive time as the other three instruments. The PI method also grossly underreported non-productive time, but the estimate of total time for PI was similar to TM because contact time was substantially overestimated by PI. Estimates of average contact time produced by PFA and ST were similar to TM, and all three of the alternate methods underestimated non-contact productive time, although not to the extent seen with non-productive time.

Figures 2, 3 and 4 are scatterplots that show the level of consistency between TM and each of the alternate methods, broken down by the three types of provider activities. Minutes measured by TM are shown on the horizontal axis, and minutes measured by the alternate method are on the vertical axis. The diagonal line represents all possible points of complete agreement between the two methods. Therefore, close grouping of points near or on the line indicates higher consistency, while dispersion or grouping of points off the line indicates lower consistency.

In Figure 2, most of the observations for contact time (the first graph) were clustered close to the line, indicating a high level of consistency between PFA and TM. Observations for non-contact productive time (the second graph) were more widely dispersed and below the line, indicating a lower level of agreement and also a tendency for the PFA method to underestimate time dedicated to these activities. The scatterplot for non-productive time showed almost no agreement with TM, because the PFA method tended to report either very low or zero non-productive time.

Figures 3 and 4 provide similar information for the PI and ST methods. PI time estimates were inconsistent with TM for all three activity categories. Patterns of non-contact productive time and non-productive time were very similar to those described above for PFA. In the case of contact time, most of the observations were above the diagonal line, indicating that providers tended to overestimate time spent with clients. The ST method was superior to the PI method in measurement of contact time, and similar to PI and PFA in measurement of non-contact productive and non-productive time.

Table 2 shows the results of a multivariate analysis that controlled for the effects of two covariates identified in the first analysis of variance, namely time of shift and clinic. The estimates in the table are averages of the absolute value of differences between the TM approach and the three other approaches. All of the differences between contact time and TM were statistically significant (p < 0.05), indicating that none of the three alternative approaches produced accurate approximations of TM times at this level of significance. Estimates of contact time produced by PFA and ST were the closest to those produced by TM.

Some of the estimated differences in Table 2 are influenced by extreme outlying values, which magnify the degree of inconsistency between the instruments. In order to reduce the impact of outliers, an alternative analytic approach was used which defined a cut-off limit of agreement between instruments of 10%, and recorded as consistent only those observations that fell within this range. Table 3 presents odds ratios that were calculated using this technique. A value greater than 1.0 shows that the instrument was more likely to produce results consistent with TM, while a value below 1.0 would indicate that the instrument was more likely to produce an inconsistent result. The PFA and ST produced estimates of contact time that were 2.7 times more likely to be consistent with TM than inconsistent, while all of the other instrument/activity combinations were much more likely to be inconsistent with TM. Odds ratios for non-productive time were not estimated because none of the instruments produced estimates within 10% of TM.

Discussion

This paper compared estimates of clinician time-use from a TM study with estimates produced by three other methods of time measurement. Two-dimensional plots of the data revealed little agreement between TM and the other methods, except in the case of contact time for PFA and the ST. Moreover, we found statistically significant differences between TM and the alternate methods for all three activity categories. We therefore conclude that, given the thresholds of agreement used in our calculations, none of the alternate methods produces estimates that are consistent with TM. We also conclude that the PI was particularly weak, because it substantially overestimated contact time while underestimating non-productive time. This finding should not be surprising, since workers have strong incentives to emphasize productive time over non-productive time. Nonetheless, the
magnitude of error in the PI estimates raises questions about the validity of the PI technique for measuring staff time.

Consistency with TM was most closely approached by estimates of contact time generated by PFA and ST. One would expect PFA and TM to yield similar estimates of contact time, since both are observational techniques that do not rely on self-reporting of time by clinicians. But the level of agreement on contact time between TM and ST is somewhat more difficult to explain, especially given that ST grossly underestimated non-productive time. One possible explanation is that providers found it easier to record time related to visits, which have discrete beginnings and endings, than to administrative tasks or other activities that occur sporadically and may require multiple attempts to complete. The constant presence of the TM observer may also have distorted clinician behaviour and reporting of non-contact time, although the magnitude or direction of error cannot be quantified.

Figure 2. Time-motion plotted against supplemented patient flow analysis

Figure 3. Time-motion plotted against provider interview
These findings may have important implications for the growing literature on productivity and costs of services.\(^4\) Firstly, it is clear that measurement and classification of non-contact time are the key issues in assessments of clinician productivity. Since none of the alternative methods produced valid estimates of non-contact time, serious questions can be raised about the findings of studies of staff productivity that rely on PI or timesheets. Of less concern, though, are the effects of using non-TM methods in studies of cost per service.

The role of time-use data in such studies is to provide a basis for allocating full costs of staff (i.e. salaries and benefits) to outputs, resulting in estimates of average cost per unit of output. A common approach is to use the percentage distribution of contact time as the allocation base, which distributes the costs of non-contact time to outputs in the same proportion as contact time. Since the entire allocation of staff time is accomplished in this one step, it is not necessary for studies of average costs to make a distinction between contact and non-contact time. But as demand for services increases, information on excess capacity at the programme level will become necessary for managers deciding how to meet higher demand.

Another important implication relates to regional and global estimates of resources needed to satisfy future demand for reproductive health services. Such projections typically are made by multiplying the unit cost of services – derived from the range of values found in the literature – by the expected demand in future time periods. But if projections are made using current average costs (which include costs of unused staff and infrastructure), they will overstate future resource requirements, because some clients could be served simply by making better use of existing capacity. The need for accurate projections of reproductive health costs is acute, and it is clear that a necessary first step is to develop techniques that accurately measure current resource use.

One potential solution could be to recommend much wider
adoption of the TM method in cost studies. But limited research budgets and scarce local expertise argue against the feasibility of widespread use of TM in many situations; indeed, such a recommendation could create a barrier limiting wider adoption of cost analysis as a routine management tool in health programmes. Future research should seek to find combinations of time measurement approaches that can produce acceptable levels of accuracy at a lower cost than TM. For example, TM could be used to measure provider time-use in clinics that are thought to be operating at full capacity (i.e. clinics where downtime is at a reasonable minimum), and these estimates could be used to define standards of contact time and administrative time for clinics that are less busy. Staff time-use could then be estimated throughout an entire clinic system. Another possibility would be to define multipliers for administrative time in a small sample of clinics, and then use contact time alone to derive estimates for all three activity categories in other clinics. Such studies could leverage the accuracy of TM, creating hybrid approaches to time measurement that generate precise time estimates at a lower cost, and also improving the accuracy of projected costs of reproductive health services.

Endnotes
1 Some recent time-use studies in the health literature have used a less labour-intensive approach known as 'random work sampling' (Knowles 1990; Finkler et al. 1993; Guarisco et al. 1994). As the name suggests, this technique collects observations on staff activities at random intervals throughout the workday. This approach is thought to cause less disruption than traditional TM techniques because staff activities are not monitored continuously. But as Finkler et al. found, work sampling is appropriate only when large numbers of observations are available, requiring many subjects and a large amount of time for observation.
2 Using actual values (positive and negative numbers) instead of absolute values would understate the differences between instruments; for example, the mean difference of an overestimate of 5 minutes (+5) and an underestimate of 5 minutes (-5) is zero, while the mean difference of the absolute values is 5.
3 The unit of analysis was the shift, the main effect of interest was the instrument pairs and the dependent variable was the difference in time spent between TM and any other instrument. The null hypothesis in this analysis was given as follows:

\[ \bar{y}_{tm} - \bar{y}_{pfa} = \bar{y}_{tm} - \bar{y}_{st} = \bar{y}_{tm} - \bar{y}_{pi} \]

where \( \bar{y}_{tm}, \bar{y}_{pfa}, \bar{y}_{st}, \) and \( \bar{y}_{pi} \) are the mean of absolute differences between TM and PFA times, TM and ST times and TM and PI differences, respectively.
4 It should be noted that these findings are based on data from three NGO clinics in one developing country; one should not automatically assume that similar results would be obtained in other settings. Two arguments can be made in favour of generalizability. Firstly, findings are consistent with expected outcomes, namely, that self-reporting methods are inherently weak because providers have incentives to emphasize productive time and understate downtime. Secondly, data from individual clinics and shifts displayed similar trends to those seen in the pooled data, suggesting that the instruments functioned consistently from clinic to clinic. Before claiming generalizability, though, the study should be replicated in different regions and programmes.

References


Acknowledgements
Funding and technical assistance was provided by the United States Agency for International Development (USAID) through Family Health International and the Population Council INOPAL II and INOPAL III projects. The views expressed herein do not necessarily reflect those of the funding agencies.

Biographies
John H Bratt holds an MA in Public Policy and an MA in Latin American History, both from Duke University. He has worked at
Family Health International (FHI) for 11 years, and his research focuses on the economics of reproductive health services.

James R Foreit, DrPH, is a Senior Associate with the Population Council, New York. He was formerly the Director of the Latin American Operations Research Project (INOPAL) and currently directs sustainability-related research and capacity-building activities for the Frontiers Project.

Pai-Lien Chen, PhD, is Senior Biostatistician at FHI. He has collaborated on epidemiological studies and clinical trials in the areas such as STD prevention studies, contraceptive efficacy and efficiency trials. His methodological research is focused in the areas of longitudinal data analysis, survival analysis and compliance analysis in clinical trials.

Caroline West has a Masters in Public Affairs and an MA in Middle Eastern Studies, both from the University of Texas at Austin. She has worked at FHI for three years, and is currently Senior Research Analyst in the Health Services Research group.

Barbara Janowitz has a PhD in Political Economy from Johns Hopkins University. She has worked at FHI for over 20 years and is currently the director of health services research. Her recent research includes studies of the cost of reproductive health services.

Teresa de Vargas is the Executive Director of CEMOPLAF, a private voluntary organization providing reproductive health services throughout Ecuador since 1974.

Correspondence: John H Bratt, Health Services Research, Family Health International, PO Box 13950, RTP, NC 27709, USA. Email: jbratt@fhi.org