THE USE OF DIRECT CLINICIAN OBSERVATION AND VIGNETTES FOR HEALTH SERVICES QUALITY EVALUATION IN DEVELOPING COUNTRIES

Kenneth L. Leonard\textsuperscript{1*} and Melkiory C. Masatu\textsuperscript{2}

\textsuperscript{1}University of Maryland, Department of Agricultural and Resource Economics

\textsuperscript{2}Centre for Educational Development in Health Arusha, Tanzania

* Author for correspondence
Acknowledgements

This work was funded by NSF Grant 00-95235, and was completed with the assistance of R. Darabe, M. Kyande, S. Masanja, H. M. Mvungi and J. Msolla. The authors are solely responsible for the data contained herein. We extend our appreciation to the Commission for Science and Technology (COSTECH) for granting permission to perform this research. The design of the vignettes benefited greatly from extensive discussion with Jishnu Das.
Abstract
This paper reports the results of a comparison between two different methods of examining quality in outpatient services in a developing country. Data from rural and urban Tanzania are used to compare the measures of quality collected by direct clinician observation (where clinicians are observed in the course of their normal consultations) and vignettes (unblind case studies with an actor). The vignettes are shown to exhibit a strong connection between the inputs provided during consultation (rational history taking, physical examination and health education) and the ability of the clinician to properly diagnose the presented illness. However, the inputs provided in vignettes are not well correlated with the inputs provided in direct clinician observation (DCO), suggesting that the inputs provided in the vignette are not well correlated with the inputs that would be provided in an actual consultation. We conclude that since vignettes do not appear to be measuring what would be provided in an actual consultation they are not a good measure of quality. Instead, we suggest that vignettes and DCO be used simultaneously. We show how the scores obtained using vignettes in conjunction with DCO can be used to improve the reliability of DCO and therefore our estimates of actual clinician quality.

Abstract word count: 203

Key words – Direct clinician observation, vignettes, health services quality, Tanzania

Full word count: 3712 (text), 4917 (full document)
Introduction

The quality of health care services is an important factor in health outcomes. Perceptions of quality can also be important to a patient’s choice of practitioner, even the choice to seek care at all. Accurate measurements of health service quality allow researchers to compare and contrast delivery systems, organizational performance and the value of training as well as to further the understanding of patient behavior. Ideally, quality is measured directly by looking for improved health outcomes. In practice, however, it is difficult to collect and use information on outcomes. Rather than measuring quality directly, many researchers measure quality indirectly, by evaluating the quality and quantity of structural and process inputs. In a developing country outpatient service context, such as that examined in this paper, structural inputs include the state of the building and grounds, the availability of clean water and toilet facilities, and the availability of key drugs, among other measures. Process inputs include politeness, consultation, dispensing of medicines, etc.

Process quality examines clinicians interacting with patients. Since the type of patient may differ from facility to facility, care must be taken to insure that process quality can be compared across facilities. This paper examines data on outpatient service process quality at health facilities in Tanzania using two different approaches: direct clinician observation (DCO) and unblind case studies with an actor (vignettes). With DCO, the researcher observes the regular consultation of actual patients. DCO allows evaluation of clinicians in an environment that is similar to the actual consultation but lacks control over the type of patients who present. Leonard, Mliga and Haile Mariam (2002) use DCO quality evaluation in Iringa Rural District of Tanzania and find that patient perceptions of quality are similar to quality as measured by that instrument. With vignettes, the researcher observes the consultation of a case study patient played by an actor. Vignettes have gained increasing popularity as a quality evaluation tool (and as a general instrument to elicit behavioral patterns in both patients and practitioners) both in developing and developed countries (Das and Hammer, 2004; De Geyndt, 1995; Epstein et al., 2001; Kalf et al., 1996; Koedoot et al., 2002; McLeod, Tamblyn, Gayton et al., 1997; Murata et al., 1992, 1994; O’Flaherty et al., 2002; Peabody et al., 1994, 1998, 2000; Sawyer et al., 1996; Tiemeier et al., 2002). Vignettes offer the researcher the chance to observe clinician behavior under carefully controlled circumstances, but may not represent the actual conditions faced by clinicians. Both instruments measure process quality
by quantifying the inputs that clinicians provide when they are consulting patients. In addition, vignettes can measure whether or not a clinician would get the correct diagnosis. In this paper, we examine and compare the results obtained from both DCO and vignettes. In particular, we intend to address the following questions:

- Is measuring inputs a valid approximation of quality?
- Do clinicians behave in the same manner for vignettes and DCO and are the results from these two instruments interchangeable?

Methods

Study Design

Cross-sectional surveys were carried out from October to December 2001, in May of 2002 and in February of 2003. Forty health facilities in the rural and urban areas of Arusha region were visited. The study population included all medical practitioners (nurses and clinicians) working in the selected health facilities. The sample includes 117 practitioners. 111 practitioners were examined with vignettes, and 96 with DCO. We were able to evaluate quality using both the DCO and vignette instruments for 90 clinicians.

Data was collected using DCO and vignettes; each method employing specific instruments developed by the research team and implemented by local medical personnel. We evaluated each facility according to the standards that would be expected of a dispensary. We looked for supplies and infrastructure that would be expected at the basic dispensary level, and examined the competence of nurses and clinicians for procedures and illnesses that could be treated at a dispensary.

Direct Clinician Observation

A member of the research team (a clinician) sat in on the regular consultations at a facility. For each consultation, the observer used a checklist of inputs --- history taking questions, physical examination procedures and health education items --- that are rational given the presenting conditions. We examine only three types of illnesses and develop specific checklists for each (fever, cough, and diarrhea). For each clinician, all presenting patients were observed, ranging from zero to 30 with an average of 13 observations per clinician. Leonard Mliga and Haile

---

1For details of sample selection and copies of the survey instruments see Leonard (2002).
Mariam (2002) find that the number of inputs provided for each consultation falls over the course of observations, suggesting that any measure of quality will have to compensate for the fact that some clinicians were observed for longer periods than others.

**Vignettes**

For vignettes, two researchers were present: a ‘patient’ (who was a female nurse) and an examiner. The examiner, after introductions, never speaks, he only observes. The ‘patient’ presented herself as a patient would. She described her symptoms and answered questions as a patient would. It was explained to the clinician that he must do physical examination by posing questions. The patient then answered the question verbally. For instance, if the clinician said, “I would take the patient’s temperature”, the ‘patient’ would say “the temperature is 38.5 degrees.”

Six vignettes were administered including one each for fever, cough and diarrhea. Since these three vignettes represent the most common illnesses presented and correspond to the three categories in DCO, our analysis is restricted to these vignettes. In addition to inputs, we graded each diagnosis given as being correct, incomplete, correct-with-extra-diagnoses or wrong.

**Methods of Analysis**

The goal of this analysis is to validate the importance of these instruments, and to compare the results between the two. We compared vignettes and DCO item by item where possible, input type by input type (where input types are history taking, physical examination and health education), by overall scores and by optimal scores.

**Validation of the DCO and vignette instruments**

We compared the levels of inputs provided to the cadre for both vignettes and DCO. In addition, we compared the level of inputs on the vignette to the diagnosis given.

**Item by Item Comparison of performance on DCO and vignettes**

There are a number of items that are included in both vignettes and DCO. These were compared item by item in order to see how often the results are the same for both instruments.
Correlation between scores for input types

Scores representing the quality of history taking, physical examination and health education were created for both vignettes and DCO. For the vignette, this score is the percentage of the total required items that were answered correct on each of the three vignettes. For DCO we used three different types of scores:

- **DCO average** The average of the percentage of total items correct across all consultations after controlling for the presenting symptom (fever, cough or diarrhea) and the age of the patient.

- **DCO maximum** The maximum score across all consultations after controlling for the presenting symptom and the age of the patient.

- **DCO random effect** The clinician constant generated from random effects regression of the score controlling for presenting symptom, the age of the patient, and the observation order allowing for random effects (correlation in errors) at the level of the clinician.¹

These DCO scores were compared to the vignette score for each type of input.

Matching overall scores

Aggregate and optimal scores for both vignette and DCO were calculated. The aggregated score was obtained by summing the scores for history taking, physical examination and health education. The optimal aggregate score was obtained from a regression using the aggregate DCO score as an independent variable and the vignette scores for history taking, physical examination and health education as dependent variables. We created three aggregate and three optimal aggregate DCO scores corresponding to the average, maximum and random effect scores.

¹ The random effect regression (or variance or error components model) assumes that each clinician contributes an unknown fixed error term to the total error. For this regression we use the observation number, the observation number squared and the observation number cubed to capture the non-linear relationship between observation and score.
Results

Validation of the DCO and vignette instruments

Table 1 shows the inputs provided as a percentage of possible inputs by cadre for both the vignette and DCO. For every category for both instruments (with the exception of health education for doctors on the vignette), higher cadres of clinicians provide the same or more inputs.

Table 2 shows the relationship between the level of inputs provided and the quality of the diagnosis for vignettes. Each cell of the table reports the number of rational input items provided as a percentage of possible input items. The percentages are compared by type and for the four different categories of diagnosis. For history taking and physical examination, providing more inputs improves the chance that the diagnosis will be correct and reduces the chance that the diagnosis will be wrong. Increasing the number of inputs provided increases the chance that a diagnosis will be incomplete rather than complete-with-extra-diagnoses. Clinicians provide more health education when they give a complete or incomplete diagnosis than when they provide a diagnosis that is correct-with-extra-diagnosis or wrong.

Figure 1 shows the relationship between the number of inputs provided for each observation on the DCO and the order of the observations. The three lines are the result of smoothing regressions and show a clear downward pattern for all three types of inputs. This pattern is important for at least two reasons. First, it is possible that clinicians are changing their behavior over time suggesting differences in the average practice, best practice and the quality of practice that patients would receive if the research team were not present. Second, if the researcher observes a clinician for only a short period, the clinician’s average score will be higher than if the researcher observes that same clinician for a long period.

Item by item comparison of performance on DCO and vignettes

There are 27 items (15 history taking and 12 physical examination) that were identical on the vignette and DCO. Table 3 looks at the relationship between these items in DCO and vignettes. For each consultation we look at whether the clinician provided one of these inputs

---

3 The lines shown in Figure 1 are the result of a local average (smoothed) regression on history taking, physical examination and health education scores after controlling for the presenting symptom and the age of the patient. Observations over 25 are set to 25 observations. We use a Gaussian smoothing kernel with a bin width of 3 observations at each of 25 points.
and compare it to whether or not he provided that input on the vignette. We report four categories for each comparison corresponding to the possible configurations of yes and no on the two instruments. The last column reports the total percentage of cases for which the answer is the same on both instruments.

Overall, clinicians provide a given input in both the vignette and DCO 26% of the time. The vignette and DCO produce the same result 63% of the time. If the clinician does the right thing on the vignette, there is a 53% chance that he will do the right thing on the DCO and a 47% chance that he will not do the right thing. The results for physical examination and history taking are broadly the same, though there is slightly more agreement for physical examination.

**Correlation between scores by input type**

We created three types of scores for the three inputs on DCO and one score for the three inputs on the vignette, leading to 9 comparisons (with 90 observations each) of which only two show significant correlation. The random effect score for physical examination is correlated with the vignette physical examination score with a correlation coefficient of 0.199 (p-value 0.06). The maximum health education score is correlated with the vignette health education score with a correlation coefficient of 0.205 (p-value 0.05). No comparisons are significantly correlated by the Spearman rank coefficient. Furthermore, we examined the correlations in scores illness condition by illness condition (fever, cough and diarrhea) and found no increase in overall correlations.4

**Matching overall scores**

Table 4 shows the correlations between the aggregate and optimal aggregate scores on the vignette and DCO. All of the aggregate scores show significant correlation between the two types of instruments. The maximum DCO has the best correlation with the vignette score, but even the random effect and average scores are correlated. By allowing the vignette score to be determined as a result of a regression, the correlation between the scores is significantly improved. The random effect DCO score is now as well correlated as the maximum DCO score.

---

4 All results are available on request.
Table 5 reports the weights assigned to each of the three vignette score types in order to obtain the three optimal aggregate scores. Reported in the cells are the coefficients (weights) and the p-values of the test that the coefficient is significantly different from zero, based on the standard errors of the coefficient. To match vignettes with either the average, random effect or maximum DCO score the greatest weight is put on physical examination in the vignette. History taking does not play a role for any of the three scores and health education is important to determine an optimal aggregate vignette score that matches with the average DCO score.

Discussion and Conclusion

Clinicians with more training provide more inputs whether measured by DCO or vignettes and clinicians who provide more inputs are also more likely to give the correct diagnosis (as seen in vignettes). Higher cadres provide higher quality because they provide more inputs. This finding validates the use of process quality as a proxy for actual quality. It does not show, however, that clinicians who provide more inputs on the vignette are more likely to give the correct diagnosis in actual practice. Quality, at least from the perspective of the patient, is not a measure of what the clinician can do, but a measure of what the clinician will do. In order to use vignettes as a proxy for actual quality we must show that clinicians who provide more inputs on the vignette also provide more inputs in DCO.

There is strong evidence that behavior on vignettes and DCO is not the same. For vignettes, doctors provide 23% more history taking and 29% more physical examination than nurses, but for DCO doctors provide only 8% and 15% more inputs respectively. The differences between cadres are smaller in actual practice (as measured by inputs provided in DCO) than they are on a test (as measured by inputs provided for the vignettes).

The item-by-item comparison of input provision for the two instruments provides additional evidence questioning the link between behavior on vignettes and DCO. At first glance, the results suggest that vignettes and DCO agree: the same result is obtained on both instruments 63% of the time. However, this is true only because clinicians are unlikely to do the right thing on either instrument. Clinicians who know what they should do (because we see that they did it on the vignette) are only slightly more likely to do the right thing on the DCO. 47% of clinicians who demonstrated that they knew what to do in a given circumstance failed to do it
for an actual patient. For these items that we can compare directly, behavior on vignettes does not predict actual behavior as observed by the DCO.

The patterns might indicate that every clinician provides half of the inputs in DCO that he provides in vignettes. If this were the case, we could not make conclusions about absolute quality using vignettes --- we could not, for example, make claims about the probability of getting a diagnosis correct. However, we should be able to make comparisons between clinicians --- we could say that if clinician A is better on vignettes than clinician B, he will also be better with actual patients. However, the comparisons between the scores obtained on the vignette and DCO suggest that this is not true.

With only two exceptions, the provision of history taking, physical examination and health education on the vignette is not correlated with the provision of these inputs in DCO. The average scores for each item are never correlated between the two instruments. Figure 1 offers a possible explanation for this result: clinicians change their behavior over the time they are being observed. This fact by itself could cause the results to be misleading. If two clinicians are the same in both ability and practice they will get the same vignette score. However, if we observe the first clinician with 10 patients and the second with 30 patients, the second clinician will have a lower average DCO score than the first even though he is not different. We believe that the decline in scores while a clinician is being observed is because clinicians alter their behavior when we first arrive (because they know they are being observed) and slowly revert to their true behavior. Other possible explanations are that clinicians tire in the course of the day, or that the types of patients change over the course of the day in a manner we fail to record. Any of these reasons imply that if we observe clinicians with different numbers of consultations we will get different scores even if they have the same quality.

The random effect score and the maximum score both attempt to circumvent this problem: the first by explicitly controlling for the order of observations and the second by giving the clinician a score based on his best performance. The correspondence between scores is better for the random effect and maximum scores but still not significant (except in the two noted cases). Controlling for the order of the observation is helpful, but there are other factors altering the behavior of clinicians between vignettes and DCO.
When we aggregate the input types to produce one composite score per clinician we find that the results improve. Even with a simple aggregation of scores, the two instruments produce results that are acceptable, although they are not strong. The aggregate vignette score is best correlated with the maximum DCO score. Vignettes measure ability and clinicians do not always practice at the level their ability. Clinicians could provide as many inputs for all patients as they do for the first few patients after we arrive, but they do not Vignettes are a better measure of the best practice of a clinician and do not reflect the actual practice of a clinician (although they are weakly correlated with actual practice.) The random effect score is a better measure of average ability because it takes into account the fact that some clinicians are observed for longer periods. This score is slightly better correlated with the vignette score, suggesting that it is important to model the effect of declining practice with observations.

The simple aggregation of the DCO and vignette scores implies that an increase in any of the three inputs on the vignette should have the same impact on the aggregate DCO score. However, a clinician who provides more physical education on the vignette than the average clinician may be very different in practice from a clinician who provides more history taking than the average clinician. In fact, the weights shown in Table 5 suggest that clinicians who do best in actual practice are those who provide high levels of physical examination and, for the average and random effect scores, those with low levels of health education on the vignette. The number of inputs provided to history taking is not an important determinant of performance in actual practice as observed by the DCO.

Using the optimal scores, it is possible to model the average actual behavior of clinicians using the vignette, despite the fact that, input type for input type, the two instruments are not correlated.

**Vignettes and DCO as Quality Evaluation Methods**

Observing clinician practice using DCO is not a perfect instrument for measuring quality in practice. However, the fact that it measures clinicians in their own environment with their actual patients makes it a better candidate to measure practice than vignettes. On the other hand, vignettes have many advantages over DCO. For example, in our research we administered a vignette on tuberculosis. In a rural clinic in Tanzania, we are unlikely to see a patient present with TB symptoms and so could not use DCO to examine the quality of
consultation for TB. In this case, the vignette offers us data that we could not otherwise obtain. Because vignettes are better correlated with best practice than actual practice, they would be useful for a study that sought to measure the impact of training on ability. However, for a research project measuring the quality of care available to the average patient, the use of vignettes by themselves would be misleading.

*Using Vignettes and Direct Clinician Observation Together*

By using vignettes and DCO simultaneously, we are able to validate the low but significant correlation between the two aggregate scores and to find the weights that allow us to create optimal aggregate scores. These optimal aggregate scores are useful in two cases. First, for many clinicians we observed very few patients. In these cases we can use optimal aggregate vignette scores (derived from weights for the whole sample) to augment our observations and increase our confidence in our quality measures. Second, we can test the behavior of clinicians for activities or types of patients that we might not otherwise be able to observe. For example, we can produce an estimate of the quality of care that a patient with TB would receive at all of the facilities that we visited, even though we did not observe any TB patients.

*Directions for Further Research*

There are many competing explanations for the lack of correlation between the scores from the three instruments. Our ongoing research focuses on the following possible explanation. Different types of clinicians might respond differently when faced with a test such as the vignette. One type of clinician might work to the best of his ability under all circumstances while another exerts extra effort under the vignette but only regular effort under the DCO. When we compare these two clinicians using vignettes we would find small differences between them but when we compare them using DCO we would find large differences. In order to test the first explanation we would need information on the characteristics of clinicians that are likely to make them respond differently to a test.
References
Kalf, AJH et al. (1996) “Variation in diagnoses: Influence of specialists’ training on selecting and ranking relevant information in geriatric case vignettes,” Social Science and Medicine, 42 (5), 705–712.
Murata et al. (1992) Prenatal Care: A Literature review and quality assessment criteria, Santa Monica: Rand Corporation.
Murata et al. (1994) “Quality Measures for Prenatal Care,” Archives of Family Medicine, 3 (1), 41–9.
Shawyer, RJ et al. (1996) “The role of clinical vignettes in rapid ethnographic research: A folk taxonomy of diarrhea in Thailand,” Social Science and Medicine, 42 (1), 111–123.
### Tables and Figures

**Table 1: Inputs provided as a percentage of possible inputs by Cadre**

<table>
<thead>
<tr>
<th>Vignette</th>
<th>Nurse</th>
<th>Clinical assistant</th>
<th>Clinical officer</th>
<th>Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clinicians</td>
<td>15</td>
<td>25</td>
<td>56</td>
<td>13</td>
</tr>
<tr>
<td>History taking</td>
<td>32%</td>
<td>41%</td>
<td>48%</td>
<td>55%</td>
</tr>
<tr>
<td>Physical examination</td>
<td>21%</td>
<td>29%</td>
<td>39%</td>
<td>48%</td>
</tr>
<tr>
<td>Health Education</td>
<td>19%</td>
<td>25%</td>
<td>27%</td>
<td>26%</td>
</tr>
<tr>
<td>Total</td>
<td>25%</td>
<td>33%</td>
<td>40%</td>
<td>45%</td>
</tr>
</tbody>
</table>

**Direct Clinician Observation (DCO)**

<table>
<thead>
<tr>
<th>Number of clinicians</th>
<th>Nurse</th>
<th>Clinical assistant</th>
<th>Clinical officer</th>
<th>Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td>History taking</td>
<td>46%</td>
<td>48%</td>
<td>52%</td>
<td>52%</td>
</tr>
<tr>
<td>Physical examination</td>
<td>32%</td>
<td>35%</td>
<td>37%</td>
<td>47%</td>
</tr>
<tr>
<td>Health Education</td>
<td>27%</td>
<td>38%</td>
<td>38%</td>
<td>51%</td>
</tr>
<tr>
<td>Total</td>
<td>36%</td>
<td>41%</td>
<td>43%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Reported is the total number of rational inputs provided as a percentage of the total number of possible rational inputs.
Table 2: Vignette Diagnosis compared to percentage of possible inputs provided

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Quality score</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>history taking</td>
<td>physical examination</td>
<td>health education</td>
</tr>
<tr>
<td>Correct</td>
<td>48%</td>
<td>39%</td>
<td>28%</td>
</tr>
<tr>
<td>Incomplete</td>
<td>45%</td>
<td>34%</td>
<td>28%</td>
</tr>
<tr>
<td>Correct-with-extra-diagnosis</td>
<td>43%</td>
<td>33%</td>
<td>17%</td>
</tr>
<tr>
<td>Wrong</td>
<td>38%</td>
<td>28%</td>
<td>19%</td>
</tr>
</tbody>
</table>

Reported is the number of rational input items provided as a percentage of possible input items categorized by the category of the diagnosis given at the end of the consultation.
Table 3: Direct clinician observation procedures compared to vignette procedures for fever, cough, and diarrhea vignettes

<table>
<thead>
<tr>
<th>Item provided for vignette?</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Same for both instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item provided for DCO?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total obs</td>
<td>6844</td>
<td>26%</td>
<td>37%</td>
<td>23%</td>
<td>14%</td>
</tr>
<tr>
<td>history taking total</td>
<td>3887</td>
<td>28%</td>
<td>33%</td>
<td>24%</td>
<td>15%</td>
</tr>
<tr>
<td>physical exam total</td>
<td>2957</td>
<td>24%</td>
<td>41%</td>
<td>22%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Shown is the correspondence for each item in each DCO consultation that corresponds to an item on a vignette. Same for both instruments implies that the item was provided for both instruments or not provided for either instrument.
Table 4: Correlation coefficients for aggregate and optimal aggregate DCO and Vignette Scores

<table>
<thead>
<tr>
<th></th>
<th>obs</th>
<th>Average</th>
<th>DCO aggregate scores</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Random Effect</td>
<td>Maximum</td>
</tr>
<tr>
<td>Aggregate Vignette</td>
<td>90</td>
<td>0.182</td>
<td>0.208</td>
<td>(0.09)</td>
<td>(0.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.09)</td>
<td></td>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Optimal Aggregate Vignette (to DCO average)</td>
<td>90</td>
<td>0.348</td>
<td></td>
<td>(0.0008)</td>
<td></td>
</tr>
<tr>
<td>Optimal Aggregate Vignette (to DCO random effect)</td>
<td>90</td>
<td>0.369</td>
<td></td>
<td></td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Optimal Aggregate Vignette (to DCO maximum)</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td>0.3592</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0005)</td>
</tr>
</tbody>
</table>

The p-value is reported in parentheses
Table 5: Coefficients and p-values for optimal aggregate regressions

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>History taking</th>
<th>Physical Examination</th>
<th>Health Education</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate average DCO score</td>
<td>0.060 (0.481)</td>
<td>0.262 (0.003)</td>
<td>-0.137 (0.083)</td>
<td>0.12</td>
</tr>
<tr>
<td>Aggregate random effect DCO score</td>
<td>0.046 (0.518)</td>
<td>0.239 (0.001)</td>
<td>-0.107 (0.105)</td>
<td>0.19</td>
</tr>
<tr>
<td>Aggregate maximum DCO score</td>
<td>0.135 (0.229)</td>
<td>0.316 (0.007)</td>
<td>-0.067 (0.518)</td>
<td>0.13</td>
</tr>
</tbody>
</table>

The p-value is reported in parentheses. Constant included but not reported. Each regression is based on 90 observations.
Figure 1: Local Average Regression Estimates of DCO Scores against observation number