Estimation of unit costs for general health services: Updated WHO-CHOICE estimates

Technical background report

Final version July 2011

Contributions

This document was prepared by a team in the WHO department of Health Systems Financing: (in alphabetical order) Georgios Gkountouras, Jeremy A. Lauer, Anderson Stanciole, Karin Stenberg, and Tessa Tan-Torres Edejer.

Funding for the project was provided by the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM). Comments on a previous version of this report were gratefully received from Eline Korenromp and Kirsi Viisainen of the GFATM.

We are grateful to the following for assistance with collection of country level data: (in alphabetical order) Baktygul Akkazieva, Fanny Esi Atta-Peters, Osmat Azzam, Aaron Beaston-Blaakman, Carmelita Canila, Mihail Ciocanu, Charles Ezenduka, Steffen Flessa, Milena Gajic-Stevanovic, Ursula Giedion, Ketevan Goginashvili, Marco Aurelio Guerrero Figueroa, Samuel Kharazyan, Miguel Linares, Mshilla Maghanga, Felix Masiye, Juan Diego Misas, Alejandro Moline, Mardiati Nadjib, Eduardo Jose Navas Coutinho, Zakariaou Njoumemi, Walaiporn Patcharanarumol, Ravindra P Rannan-Eliya, Mohamed Samai, Mariyam Sarfraz, Bill Stomfay, Siok Swan Tan, Tsolmongerel Tsilaajav and Kirsi Vitikainen.

Comments on the report should be forwarded to Karin Stenberg (stenbergk@who.int).

Disclaimer

All views expressed in this report are those of the authors. The authors have no conflict of interest to declare.

1. Background

1.1. Cost information for policy analysis

Policy makers rely on high quality information to make informed decisions. The need for information on which to base resource allocation is particularly pertinent to the health sector since health expenditure constitutes a large component of the annual government budget in many countries. Information on the cost of health services can serve various purposes including budgeting, cost-effectiveness or cost-benefit analysis, or the comparative assessment of efficiency in different settings. Information on the cost effectiveness or budgetary implications of different strategies and interventions is important for countries. To inform these decisions, accurate estimates on both health effects and costs is essential.

Recognizing the need for country-specific and facility-level information on costs, WHO has collated data from countries around the world as part of its WHO-CHOICE (CHOosing Interventions that are Cost Effective) project since early 2000.[1] In 2000, WHO-CHOICE estimated the costs of health service utilization (inpatient days and outpatient visits) by country based on multiple regression analysis done on data from 72 countries for various years between 1980 and 2000 (about 90% of the data was post-1990) for a total of 2415 country years of observations.[2] These estimates are made publicly available (http://www.who.int/choice/country/en/index.html) and have been used by researchers, academics and analysts at both global and country level. One example is the project on Disease Control Priorities in Developing Countries (DCPP).[3] At country level, numerous projects performed in partnership with ministries of health have used the estimates.

The costs of health service utilization form an important component of the overall per-person costs of health services; however, disease-specific control programmes however do not typically collect or incorporate them in their reports. With priority programmes supported by global health initiatives scaling up and increasing demands on the health system, information on the costs of service delivery can be used as an indication of the need for health system support. The production of up-to-date country estimates of the costs of service utilization is therefore important to inform programme-level assessments of the budgetary needs as well as the cost effectiveness of HIV, TB and malaria services.

As countries' use of technology evolves, so do their costs. The need to update the 2000 estimates was prompted by the expectation that inputs (including technology), prices and production efficiency all could have changed in the last decade. The 2000 estimates were updated in year 2008 to reflect 2005 price levels by substituting new input values for the independent variables in the original 2000 regression analysis. However a need to engage in a new round of data collection to assess to what extent changes in policy practices and technology may have also affected costs was recognized.

In 2008, a decision was therefore made to engage in a new round of data collection. The updated analysis was funded by the GFATM. This report provides an overview of the methodology and findings.

1.2. Objective

To estimate current unit costs of outpatient visits and inpatient days for a variety of facility types in low- middle- and high-income countries.

[1] Hutubessy, R. C., et al (2003).

[2] Ibid., Adam, T.,et al (2003)

[3] Disease Control Priorities in Developing Countries (2nd Edition) , (2006)

1.3. Methodological approach

An outline is provided here with more details in sections 2-4 below.

1.3.1. Literature review

A review of both publications and the grey literature on the costs of health service utilization, especially in the context of developing countries, was undertaken.[4] The objective of the review was to assist in the selection of relevant variables and methods that should be taken into account during data collection and analysis, and allowed for the identification of individuals and institutions that have collected cost data.

1.3.2. Assembly of unit cost data sets from a sample of countries from all regions

Cost data was collected through various mechanisms. Firstly, authors identified through the literature review were contacted in an attempt to access the raw data used for publications. In addition a public call was issued to identify owners of unit-cost datasets. A standardized template was prepared to allow for data collection for a specific set of variables. Finally, attempts were made to access publicly available data from ministry of health websites.

1.3.3. Database management

The data collected were aggregated and analysed in terms of summary and descriptive statistics. This process assisted in clarifying the availability of data from countries and their quality, and also informed on the scope of variables available for use in the econometric analysis. Previous experience suggested that some variables are more capable of influencing the outcome of the analysis. For instance, it was clear that the level of the health facility (i.e. primary, secondary or tertiary) was a main cost driver. Other variables, such as the proportion of emergency admissions, did not prove to be as important. Experimentation with different variables showed that a restricted list was preferable.

1.3.4 Econometric analysis to estimate unit costs of outpatient visits and inpatient days for all countries

Examination of the 2007 dataset and comparison with data previously collected (i.e. for the 2000 analysis) revealed that the new data had very different characteristics. Some of these differences were due to differing definitions of similar variables; others were attributable to altogether different variables that were collected, as a result of differing collection methods and instruments. Moreover, there was strong evidence of statistical heterogeneity in variables that could be directly compared. As a result the two datasets were not pooled, and regression analysis was performed using only the new dataset.

[4] Stanciole AE, et al.,(2009).

2. Data collection process

2.1. Locating and contacting data providers

Experience from the previous exercise suggested that carrying out 'bottom-up' facility-level estimation would be expensive, as well as probably not very efficient when applied to the required number of countries. An alternative strategy was therefore employed.

Cost data were collected through the following three mechanisms:
(i) contacting authors identified in the literature review;
(ii) sending out a public call for proposals; and
(iii) accessing publicly available data.

Authors identified as owners of data sets that could be considered for inclusion in the study were contacted. However, only two published studies fit the criteria and were subsequently included in the project.

A call for proposals was sent to individuals and institutions that potentially had data available. The objective was to collect data from research that had already been completed, rather than commissioning new data collection.

Respondents to the call were sent a scoping questionnaire to assess the type of data available (see Annex 1). Over 60 proposals were submitted and reviewed. If, based on the scoping questionnaire, the quality and comprehensiveness of the data were deemed suitable then the proposal was accepted and the data owners were requested to make their data available to WHO. They were asked to identify the time that would be required to sort and present their data in line with the required standards, and if further analysis was deemed necessary they were offered a contract to perform the work. Contractors were subsequently sent a data-collection template and other necessary documentation.

We also actively contacted potential contractors whom we knew were likely to have facility-cost data available, based on their participation in the previous (2000) analysis or from collaboration on other projects.

Based on the responses obtained from the call for proposals, and considering geographic and income-level representation, a total of 30 proposals for further work were accepted, of which three were subsequently cancelled do to non-compliance with the contract.

In addition, the World Wide Web was searched for public databases, using the key words “ministry of health “ and “database” in various languages. The search included websites from ministries of public health from various countries (Australia, USA, UK, Spain, France, Germany, and Portugal). Only data from USA was extracted directly from the website and could be included in the final data set. In addition, Australian data were obtained from direct contact with the National Hospital Cost Data Collection (NHCDC) project. The reasons for not including other publicly available data sets were because they did not present the requested data on a facility level basis. Some of the datasets provided data on total facility costs but with no breakdown into inpatient and outpatient costs as required for analysis (e.g. Spain), or presented costs at a different level than that of a hospital (e.g. UK, Portugal). For other countries it proved to be impossible to extract or interpret the data without direct assistance from ministry staff (e.g. France), which idea was abandoned due to the inability to contact the appropriate staff or due to time constraints or both. In some countries cost data by facility was not published in the national database due to legal barriers (e.g. Germany).

In sum, 31 data sets were included of which 2 were identified through the literature review (i.e. Philippines, Burkina Faso), one was located and downloaded from the Web (i.e. USA), and one was located through actively contacting known data holders in the country (i.e. Australia). Table 2 gives a summary. Since two data sets referred to the same country (Philippines), a total of 30 countries were represented.

2.2. Survey questionnaire used to screen data providers

The survey questionnaire was developed drawing upon experience from the previous round of data collection, as well as the findings from the literature review. Contractors used standardized instruments to provide summary information on a sample of health facilities: total annual expenditures, total outputs (inpatient days and outpatient visits) and other key variables such as staff salary costs, occupancy rates, and other explanatory variables thought to be plausible or found to be explanatory in previous studies.

The survey documentation included definitions of all variables. The survey was designed to elicit cost information relevant to a classification comprising four types of facility as shown in Table 1.

Table 1: Definition of facility types

Facility type	Description
Level 1	Health centre with outpatient services only
Level 2	Health centre with limited number of day beds (mainly maternity)
Level 3	Hospital primarily for simple cases (e.g. district hospital)
Level 4	Specialist hospital (e.g. referral hospitals)
Level 5	Teaching hospital (= level 4 hospital with teaching component)

2.3. Follow-up and quality control

Data collection took place during a 12-month period from July 2009 to July 2010. Respondents were screened for data quality and consistency with the requested data format. Intensive follow up was typically required in order to ensure that the data provided matched the requested criteria. Discussions were held on methods to enhance the quality and representativeness of the data when they were incomplete or not in conformity with the specification. Many contractees experienced delays in providing the data due to change in personnel at the ministry of health, or due to problems allocating costs to inpatient and outpatient activities.

3. Data

3. 1 Sources of data

As described above data were obtained from three sources:

1) Where the sample was large enough and information was provided on all relevant variables, reports of costing studies identified in the literature review.
2) Providers who agreed to enter into a contract.
3) Data publicly available from national health authorities.

3. 2 Data categories

Data collection templates requested information about general facility characteristics (name, affiliation, urban/rural location, facility type, financing type (e.g. private or public)), on costs, including total costs, costs by category of input (e.g. medicines, salaries), costs by activity (e.g. outpatient and inpatient care), as well as on outpatient and inpatient unit costs. Information on activities (e.g. total hospital admissions, number of procedures, average length of stay).

3. 3 Total number of countries included

The final sample included data from 30 countries representing 13 WHO epidemiologic regions. The names of the countries are shown in Table 2. The majority of data was obtained through contracts with local data providers. Only one usable dataset was identified from the literature review,[5] as data from most of the publications reviewed was too limited in terms of the variables reported.

With regards to public data sets, the only ones included in our analysis were from the United States and Australia (the latter requiring further contact with data managers to ensure usability).

Table 2: Distribution of data points among countries and WHO epidemiologic regions

WHO Region	Country	Source of data	Health centres	Hospitals	Total	World Bank income classification	% of total sample
AfrD	Benin	Public call		29	29	Low-income	0.0%
	Burkina Faso	Literature review	25	1	26	Low-income	0.0%
	Cameroon	Public call	39	6	45	Lower-middle-income	0.0%
	Ghana	Public call	158	41	199	Low-income	1.0%
	Nigeria	Public call	13	17	30	Low-income	0.0%
	Sierra Leone	Public call	956	42	998	Low-income	5.0%
AfrE	Rwanda	Public call	36	7	43	Low-income	0.0%
	Uganda	Public call	9	3	12	Low-income	0.0%
	Zambia	Public call	127	17	144	Low-income	1.0%

[5] Tsilaajav, T., Costing study for selected hospitals in the Philippines, 2009.

AmrA	USA	Website	-	4,938	4,938	High-income	26.0%
AmrB	Brazil	Public call	6,725	2,162	8,887	Upper-middle-income	47.0%
	Colombia	Public call	771	154	925	Lower-middle-income	5.0%
	Dominican Republic	Public call	41	20	61	Lower-middle-income	0.0%
	El Salvador	Public call	-	270	270	Lower-middle-income	1.0%
AmrD	Ecuador	Public call	36	26	62	Lower-middle-income	0.0%
EmrB	Lebanon	Public call	-	81	81	Upper-middle-income	0.0%
EmrD	Pakistan	Public call	25	10	35	Low-income	0.0%
EurA	Finland	Public call	-	37	37	High-income	0.0%
	Netherlands	Public call	-	37	37	High-income	0.0%
EurB	Armenia	Public call	32	106	138	Lower-middle-income	1.0%
	Georgia	Public call	247	101	348	Lower-middle-income	2.0%
	Kyrgyzstan	Public call	9	54	63	Low-income	0.0%
	Serbia	Public call	60	46	106	Upper-middle-income	1.0%
EurC	Republic of Moldova	Public call	84	63	147	Lower-middle-income	1.0%
SearB	Indonesia	Public call	17	22	39	Lower-middle-income	0.0%
	Sri Lanka	Public call	17	52	69	Lower-middle-income	0.0%
	Thailand	Public call	199	857	1,056	Lower-middle-income	6.0%
WprA	Australia	Data provider contacted directly		76	76	High-income	0.0%
WprB	Mongolia	Public call	23	16	39	Lower-middle-income	0.0%
	Philippines (*)	(a) Public call /, and (b) Literature review	23	45	68	Lower-middle-income	0.0%
Total			**9,672**	**9,336**	**19,008**		

(*) Note: two data sets were obtained from Philippines and combined. These are shown here as one category.

While the current (2007) database includes fewer countries than in the previous (2000) round (30 compared to 80), the previous (2000) database included mainly small datasets. Considering just datasets with more than 10 facilities, the previous (2000) database included only 33 countries (Table 3). The updated database can thus be considered similarly representative of between-country variation and more representative of the within-country variation (due to a higher average number of facilities within countries).

Adam and Evans (2006) noted considerable variation between reported costs within countries and emphasized that a sufficiently large number of facilities within a country is required to ensure representativeness.

Table 3. Number of countries included in unit cost database: original and updated analysis

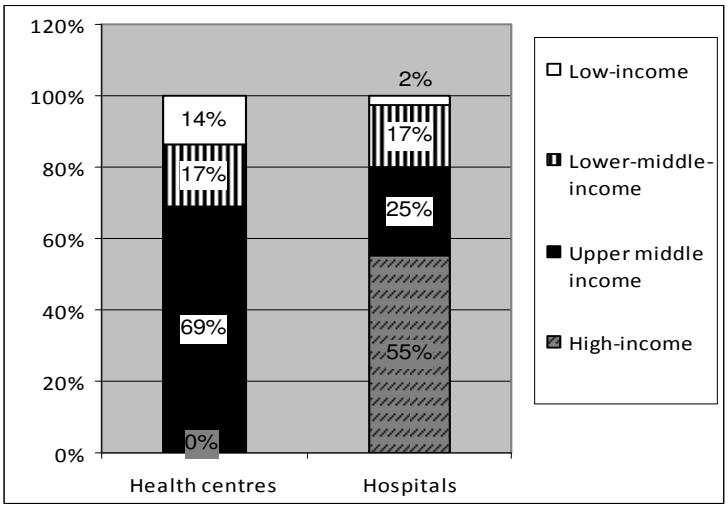
	2000 estimations				2007 estimations			
Country classification	Number of countries	% of countries in sample	of which countries with <10 facilities	of which countries with >10 facilities	Number of countries	% of countries in sample	of which countries with <10 facilities	of which countries with >10 facilities
High-income	29	36%	18	11	4	13%	0	4
Upper-middle-income	12	15%	8	4	3	10%	0	3
Lower-middle-income	21	26%	10	11	13	43%	0	13
Low-income	18	23%	11	7	10	33%	0	10
Total	80	100%	47	33	30	100%	0	30

3. 4 Distribution of countries by income group

The main difference between the current (2007) and the previous round of analysis is the relatively fewer number of high-income countries included. Significant challenges were encountered in accessing data from high-income (e.g. OECD) countries as mentioned above. These include legal barriers and a high level of aggregation in publicly available data sets.

Figure 1 shows that high income countries made up 56% of the data for hospitals but 0% of the health centre level data. As the 'health centre level' in high-income countries is effectively equivalent to doctor's surgeries/clinics, it is doubtful that a meaningful comparison of 'health centres' could be effected across high- and low-income countries. In any case, the effect of this feature of the data on the comparability of cost estimates for similar services (e.g. outpatient visits) across both hospitals and health centres is unknown.

Figure 1: Distribution of data points by income level (2007 WHO-CHOICE dataset)



3. 5 Size of database in terms of number of data points

The data collection efforts have resulted in a significantly expanded dataset compared to the original database. The overall sample is 6 times larger, with five times as many data points for hospitals and 9 times as many data points for Health Centres. The increase is the greatest for upper middle income countries.

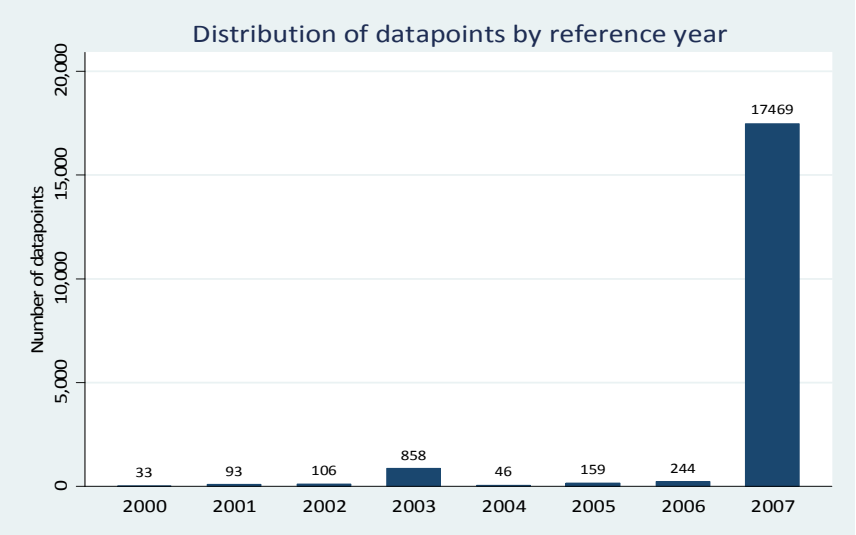
Table 4. Number of data points included in unit cost database: original and updated analysis

	Original (2000) analysis				Updated (2007) analysis			
World Bank country classification	**Health centre**	**Hospital**	**Facilities**	**% of sample**	**Health centre**	**Hospital**	**Facilities**	**% of sample**
High-income	101	807	908	32%	0	5088	5088	27%
Upper-middle-income	124	76	200	7%	6785	2289	9074	48%
Lower-middle-income	619	769	1388	49%	1529	1738	3267	17%
Low-income	232	120	352	12%	1358	221	1579	8%
Total	1076	1772	2848	100%	9852	9208	19060	100%

3. 6 Database by reference year

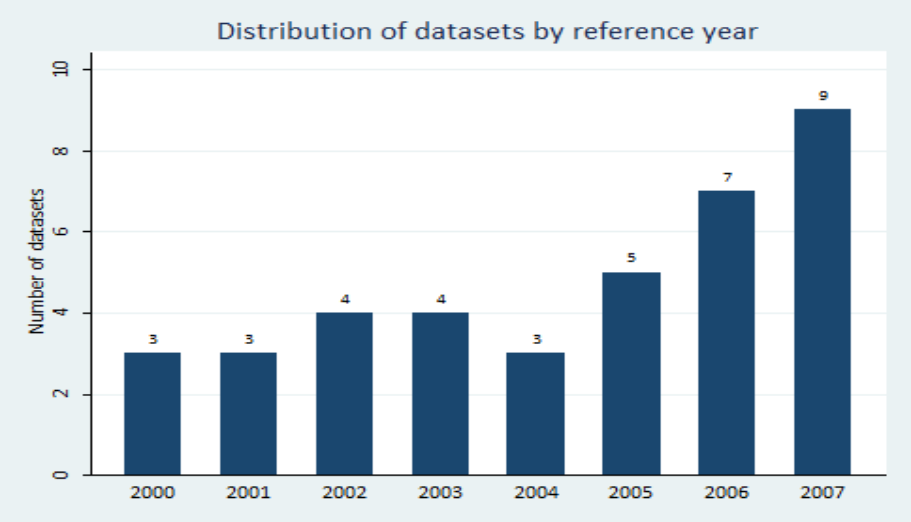
The 2007 dataset represents recent data with 92% of estimates referring to year 2007 (Figures 2 and 3).

Figure 2. Distribution of data points by reference year [6]



[6] Data point: refers to an estimate belonging to a facility in a country in a specific year. Data set: refers to a group of facilities in a country in a specific year. Data sets consist of data points. For example, Finland consists of two data sets, with one set containing 36 observations (for year 2007) and the other set including one observation (for year 2002).

Figure 3. Distribution of year-specific data sets by reference year



3. 7 Challenges in data collection

In this section we provide an overview of some of the challenges encountered in data collection and how they affected the analysis.

3.7.1 Datasets provided by countries

An overall challenge is that many datasets had key variables missing. This was especially true for the breakdown of costs into cost components such as salaries, drugs, lab tests, and other. Some contractors provided information on 100% of requested variables whereas other provided information on less than 50% of requested variables. The average completion rate for the entire sample was 56%. The lack of data influenced which variables could be used for the regressions. Table 5a (inpatient costs) and Table 5b (outpatient costs) illustrate the extent of missing data for the variables chosen as determinants in the final regressions.

Table 5a. Variables collected from countries related to inpatient unit costs

Variable	Description	No. observations*	No. missing
inpatient	Inpatient unit cost expressed in 2007 Int. $	7942	1394
gdp	GDP per capita (PPP)	9336	0
pctwardbeds	Occupancy rate	9169	167
alos	Average length of stay	8868	468
admissions	Total inpatient admissions per year	4802	4534
HOSP1	Dummy variable for level 3 facilities	9336	0
Dteach	Dummy variable for teaching hospitals	9173	163
DBrazil	Dummy variable for observations of Brazil	9336	0
public/private/NGO	Dummy variable for public or private level hospitals	9336	0

*Only considers facility levels 3 and 4.

Table 5b. Variables collected from countries related to outpatient unit costs

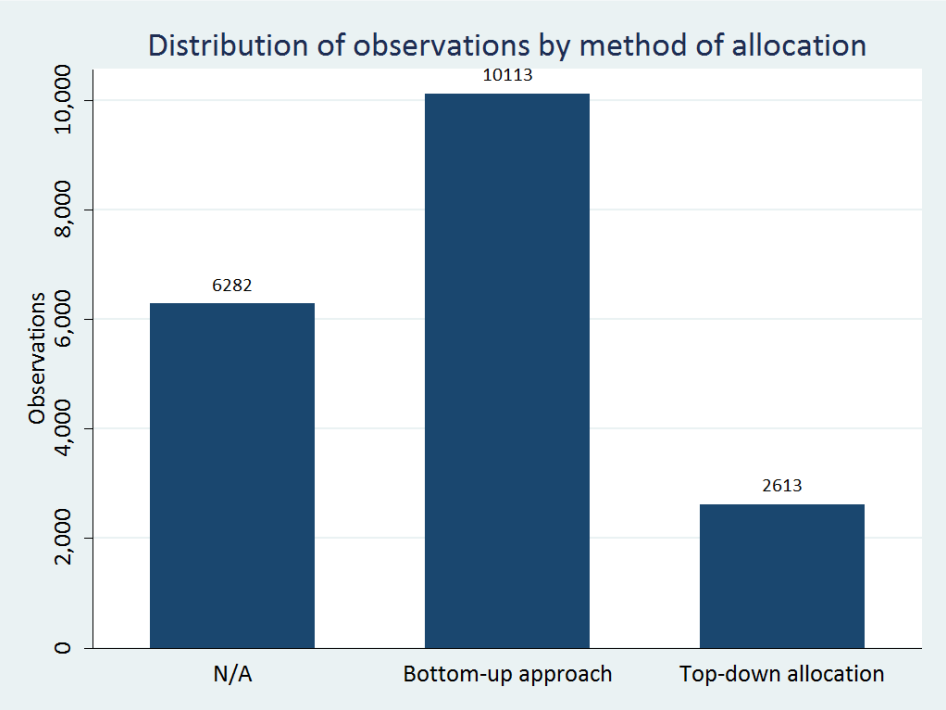
Variable	Description	No. observations	No. missing
outpatient	Outpatient unit cost expressed in 2007 Int. $	14479	4529
gdp	GDP per capita (PPP)	19008	0
visits	Total outpatient visits per year	14573	4435
viz2	Visits per provider per day (nurses, GPs)	9344	9664
urban	Dummy variable for urban location of the facility	19008	0
Public/private/NGO	Dummy variable for public or private level hospitals	19008	0
HC2	Dummy variable for level 2 facilities	19008	0
HOSP1	Dummy variable for level 3 facilities	19008	0
HOSP2	Dummy variable for level 4 facilities	19008	0
DColombia	Dummy variable for observations of Colombia	19008	0
DBrazil	Dummy variable for observations of Brazil	19008	0
Dbrazil3	Dummy variable type 3 facilities of Brazil	19008	4529

Limited information was provided on the breakdown of costs into components (e.g. salaries, drugs, lab tests and other costs). Even when data on the breakdown of costs were provided questions remained with regards to their completeness, for example some countries reported drugs as <1% of total costs. This problem impeded the development of a one-step regression model with drug costs as an independent variable (see *Methods for regression analysis*, below).

Second, different methods were used across countries. For example, contractors were requested to provide information on how their estimates of unit costs were estimated - either using a bottom-up approach or a top-down approach. A total of 53% of costs were reported to be estimated using bottom-up methods and 14% were reported as derived using a top-down approach (for 33%, no information was provided on estimation method). Moreover, within these categories, methods may still differ regarding the precise list of components included within costs.

A third limitation is that for some countries the data provided referred to patient charges and not full economic costs (e.g. Benin and Georgia).

Figure 4: Methods of cost allocation reported within the sample



3.7.2 Publicly available datasets

Review of publicly available datasets revealed that most did not have information on all required variables. Many countries with publicly available data only provide information on process but not on costs by hospital or by health facility. When cost information was available, it was often not in the format required for analysis.[7]

3.8 Selecting data for inclusion in the regression analysis

As shown above in section 3.7.1 significant data were missing. The total number of variables used for inpatient cost estimation was 3407,[8] whereas for outpatient costs 9028 data points were used.[9] Tables 6A and 6B provide a description of the variables for the data on which the regressions are based. Figures 5 and 6 provide an overview of the distribution of data points used in final regression models, by income level

[7] For example, United Kingdom had information organized by NHS trust regions and the unit costs that were provided were on average national level.

[8] This is greater than the number of observations used in the 2000 analysis by Adam et al. (2003) where the full database covered 2173 country-years of observations, and where only 1171 data points (country-years) were used for the final regression.

[9] The number of observations used here compared to the previous analysis cannot be directly compared as there were two distinct models for outpatient costs in the previous (2000) analysis.

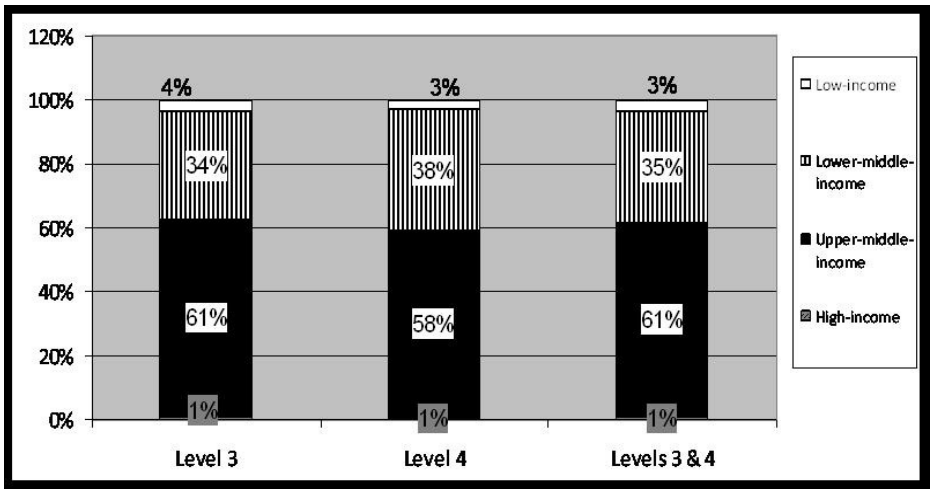
Table 6A. Variables for the inpatient unit cost regression and their distribution

Variable	Description	N	Mean	SD	Median
inpatient	Inpatient unit cost expressed in 2007 Int. US$	3407	181.59	241.21	90.14
gdp	GDP cap(ppp)	3407	8788.86	3729.88	9566.66
pctwardbeds	Occupied ward beds	3407	0.23	0.33	0.04
alos	Average length of stay	3407	4.16	2.33	3.41
admissions	Inpatient admissions	3407	4700.00	7416.81	2509.00
HOSP1	Dummy variable for level 3 facilities	3407	0.80	0.40	1
Dteach	Dummy variable for teaching facilities	3407	0.09	0.29	0
DBrazil	Dummy variable for observations of Brazil	3407	0.58	0.49	1
public	Dummy variable for public level hospitals	3407	0.58	0.49	1
private	Dummy variable for private level hospitals	3407	0.16	0.36	0

Table 6B. Variables included in outpatient unit cost regression and their distribution

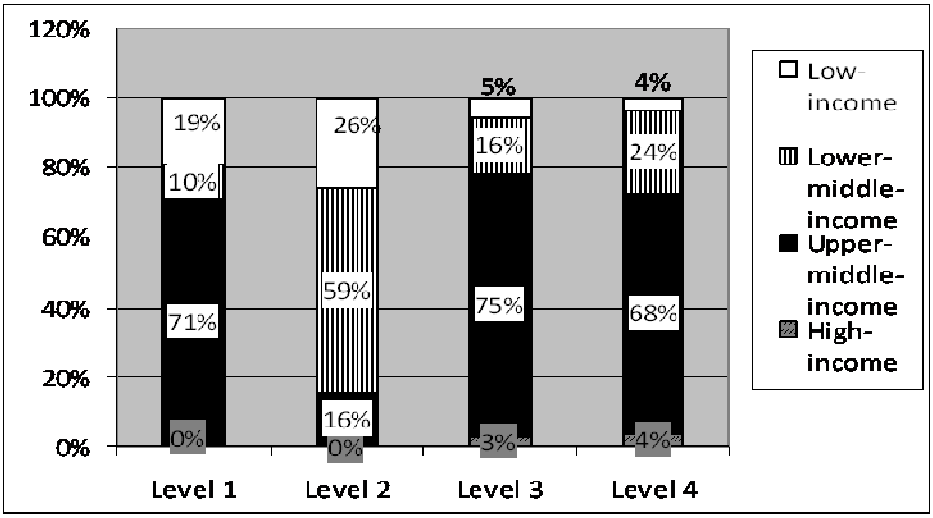
variable	Description	N	Mean	SD	Median
outpatient	Outpatient unit cost expressed in 2007 Int. US$	9028	15.56	44.02	5.50
gdp	GDP cap(ppp)	9028	7947.64	4618.50	9566.66
visits	Outpatient visits	9028	62924.53	118929.70	19414.50
viz2	Visits per provider per day (nurses, GPs)	9028	5.71	10.90	2.62
urban	Dummy variable for urban location of the facility	9028	0.21	0.41	0
public	Dummy variable for public level hospitals	9028	0.75	0.44	1
private	Dummy variable for private level hospitals	9028	0.12	0.32	0
HC2	Dummy variable for level 2 facilities	9028	0.15	0.35	0
HOSP1	Dummy variable for level 3 facilities	9028	0.25	0.43	0
HOSP2	Dummy variable for level 4 facilities	9028	0.08	0.26	0
DColombia	Dummy variable for observations of Colombia	9028	0.10	0.30	0
DBrazil	Dummy variable for observations of Brazil	9028	0.62	0.49	1
Dbrazil	Dummy variable type 3 facilities of Brazil	9028	0.18	0.39	0

Figure 5: Distribution of data points used in final inpatient unit cost regression, by income level



Inpatient unit cost data for levels 3 and 4 is significantly lower than the share indicated in Figure 1, due to the elimination of in-patient unit costs data from the USA dataset.

Figure 6: Distribution of data points used in final outpatient unit cost regression, by income level



The regression analysis model for level 1 data has a somewhat lower dependence on data from upper middle income countries compared to the total sample of data collected (refer to Figure 1). In general outpatient cost data from high income countries is limited.

4. Methods for regression analysis

4.1. Deriving the regression models: approach

The approach adopted here derives from the economic literature on 'hybrid cost functions' (Pauly, 1986). In this construct, a log-cost function such as that faced by a health facility is assumed to depend on:

- A log-additive vector of input prices, and
- An unknown function of:
    - a set of output indicators
    - a set of variables indicating facility type.

Although the function of output indicators and variables indicating facility type is held in general to be of unknown form, in practice, a log-additive specification is usually adopted. Insofar as the specification of the log-cost function is log-linear, the model is consistent with a Cobb-Douglas production function.

In order to estimate the outpatient and inpatient unit costs, various logarithmic models were tried and tested. The models also drew upon the regressions used for the previous work estimation of inpatient unit costs carried out at (Adam et al. 2003, Adam & Evans 2006, Adam et al. 2008). Variables were chosen based on the following criteria:

- the variable is a known determinant of unit cost,
- measurement data for the variable are readily available,
- the variable performs well in regression models.

We tried a number of different log-log OLS regression specifications before arriving at a final model choice. Each model was assessed based on fit statistics and the significance of the explanatory variables, as well as on internal and external validity criteria. The model specifications we report are robust in the sense that, in any alternative model specification tested, the coefficients of most variables included in the reported specifications were strongly significant; conversely, the coefficients of the variables not included in the reported model specifications were generally not significant. Other standard tests of statistical validity were also assessed.

In addition, and, for determining the final models for prediction[10] purposes, we also assessed whether the signs of the estimated coefficient were in line with predictions from economic theory and previous findings reported in the literature. Finally, for the purposes of predicting unit costs, it was moreover essential to employ as predictors those variables for which values could be assigned, in the sense that the predictions of unit costs are of little use if they rely on specialized knowledge about country situations beyond that which could be considered generally available.

Another major challenge for model selection was dealing with the issue of drug costs. One of the goals of the current project, common with the previous work done at WHO, is to be able to predict costs in a facility of a given type in an arbitrary country. Moreover, predicted costs have always been understood as costs net of the cost of drugs, food and labs (more specifically, the estimates of unit costs are understood to include the 'hotel' costs of the facility visit as well as the medical-staff salary costs) In the previous (2000) work on unit cost estimation done at WHO, a dummy variable was coded to take the value of '1' when a facility reported drug costs, and to take the value of '0' otherwise, both in the regression

[10] 'Estimation of unit costs' using the regression models is called 'prediction', in line with common statistical usage.

equation used for the estimation of the unit cost of inpatient days and in the equation for the estimation of the unit cost of outpatient visits. By use of this simple device, it was confirmed that facilities reporting non-zero drug costs had, on average, a higher reported unit cost than other facilities. Thus, for the purposes of prediction of facility-specific unit costs, the dummy variable in the estimated regression equation was set to '0', and the resulting predicted facility-specific unit cost could be interpreted as a 'unit cost net of drugs'.

In the latest data collection exercise, however, it was found, first of all, that most facilities did not report drug costs in spite of detailed instructions requesting this information. Secondly, it appeared that the validity of the values that were reported was questionable, largely because drug costs were found to vary from a minuscule percentage of total facility costs to a multiple many times greater than total facility costs. Moreover, it proved impossible, for those facilities offering both inpatient and outpatient care, to apportion drug costs between the two respective cost centers in any meaningful way. Thus, it was hard to avoid the conclusion that no single, common method had been used across countries to classify the cost category pertaining to drugs. Finally, although numerous specifications were tried, as well as imputation of missing values, it was found impossible to find a specification for drug costs that performed sensibly in either the regressions or with respect to simple descriptive and summary statistics.

In the light of this finding, a method was implemented, as follows, to adjust facility-specific predicted unit costs for the drug cost component:

1. Using the previously published Adam et al. models, to predict inpatient unit costs and outpatient unit costs with the dummy variable for 'drug costs' set equal to 1;
2. Using the previously published Adam et al. models, to predict inpatient unit costs and outpatient unit costs with the dummy variable for 'drug costs' set equal to 0;
3. On the basis of these predictions, to calculate 'drugs costs' as a proportion of inpatient unit costs and outpatient unit costs based on the Adam et al. models;
4. Finally, to subtract this same proportion from our own estimates of inpatient unit costs and outpatient unit costs, resulting in estimates of facility-specific unit costs net of drug costs.

A similar approach was adopted for separating out food costs, i.e. the previous Adam et al. models were used to predict values with and without food costs, and the proportional food cost was subtracted in order to arrive at a net estimate for the 'hotel' and medical-salary components of costs.

The importance of including variables to measure the capacity and size of facilities was recognized early on. For the inpatient model, capacity was measured by the percentage of beds occupied. For the outpatient model, we used the number of visits per provider per day. However, we could not use the number of beds to measure facility size for the inpatient model, due to its high colinearity with the percentage of beds occupied. Thus we used the number of admissions as a variable to indicate size. For the outpatient model, the number of visits per year was used as an indicator of size.

4.2 Model specification and results

OLS assumptions are assumed to apply to all models.

4.2.1 Inpatient unit cost regression specification and results

The model estimating facility-specific inpatient unit costs uses as explanatory variables:

- the logarithms of:

  - GDP per capita,
  - occupancy rate,
  - total inpatient admissions,
  - average length of stay,
- and dummy variables to indicate:
  - whether the hospital is used for the teaching of health workers,
  - whether the hospital is a third or fourth level facility,
  - whether the hospital is financed by public, private or NGO funds, and
  - whether the hospital is located in Brazil.

As a detailed vector of input prices is not available, it is proxied here by the variable for GDP per capita. Furthermore, similar to the rationale of Adam et al. in their previous analyses, GDP per capita also serves as a proxy for the level of technology. Regarding the set of output indicators that such models typically employ, occupancy rate, total inpatient admissions and average length of stay together provide a measure of capacity utilization while controlling for facility size. An indicator of the level of the facility and for whether the hospital is employed for the teaching health workers together provide an indirect (i.e. 'proxy') measure of case mix while controlling for facility size. A dummy variable for whether the facility is located in Brazil is included to account for the fact that many observations in the sample were from Brazil and, even after controlling for GDP per capita, it is reasonable to expect that hospitals in Brazil face a systematically different cost structure than similar facilities in other locations.

The algebraic form of the inpatient unit cost model is:

$$\ln IUC_i = a_0 + a_i \sum_{i=1}^{n} \ln X_i + e_i, i = 1...n$$

However, back transformation of predicted log unit costs (iuc_i) gives the median and not the mean of the distribution. Therefore, a bias correction technique is needed. Following Adam & Evans (2003), we use the 'smearing' method proposed by Duan (1983). The method is non-parametric, because it does not require regression errors to have any specific distribution. The smearing correction factor was estimated following three steps: i) Estimation of regression residuals, r_i. ii) Exponentiation of regression residuals to the power *e*, $\exp(r_i)$. iii) Averaging of the exponentiated residuals $1/n * \sum \exp(r_i)$. The correction factor is then used to multiply the back-transformed estimated log unit costs (i.e. $\exp(c_i)$).

Robust estimation methods were used (i.e. the *Stata* command "robust"), in order to control for the effect on the estimate of standard errors caused by 'clustering' (i.e. the inclusion of multiple observations per country).

Outliers were removed from the sample on which the regressions were based. Annex 3 provides more detail on the boundaries imposed on outliers.

Table 7 shows the final regression model for inpatient unit costs. The model differs in terms of explanatory variables from that of Adam & Evans (2003) – see Annex 4 for an overview of the regression models previously published. For example, we have not included variables for food and drug costs because, as explained above, the data did not allow such a specification. On the other hand we included other variables that have been reported in hospital cost function estimation literature.[11]

Table 7. Regression output for inpatient unit cost expressed in 2007 Int. $

	Regression coefficient	**95% confidence interval**
Natural log of GDP per capita (PPP)	1.192***	[1.111,1.272]
Natural log of occupancy rate	-0.0201**	[-0.0340,-0.00623]
Natural logarithm of ALOS	-0.600***	[-0.649,-0.550]
Natural logarithm of total inpatient admissions	0.0252*	[0.00471,0.0457]
Dummy variable for level 3 facilities	-0.204***	[-0.275,-0.132]
Dummy variable for teaching hospitals	0.257***	[0.163,0.351]
Dummy variable for public level hospitals	-0.144***	[-0.182,-0.107]
Dummy variable for private level hospitals	0.110***	[0.0710,0.148]
Dummy variable for observations in Brazil	-1.638***	[-1.694,-1.583]
Constant	-4.277***	[-5.035,-3.519]
Observations	3407	
R^2	0.760	
Adjusted R^2	0.760	
AIC	3988.2	
BIC	4049.5	
F_stat	1070.3	
Prob	0	
RESET	69.18	
RESET_P_value	2.02e-43	
Log_likelihood	-1984.1	
CF	1.054	
VIF	1.721	

*** *p* < 0.05, ** *p* < 0.01, *** *p* < 0.001**

Note Only type 3 and type 4 facilities are included in this regression model.

4.2.2 Outpatient unit cost regressions specification and results

For the estimation of facility-specific outpatient unit costs, we investigated several approaches, and the one that performed the best according to the criteria defined above was a regression model estimating outpatient unit costs for a pooled sample of health centers and hospitals. We also investigated an approach similar to that found in Adam & Evans (2006) and Adam et al. (2008) (i.e. indirect estimation), but the pooled approach was judged to be preferable, as it is a direct estimation technique and it integrates information from all facilities and captures their differences in unit costs.

The algebraic form of the outpatient unit cost model is:

$$\ln OUC_i = a_0 + a_i \cdot \sum_{i=1}^{n} X_i + e_i, i = 1...n$$

[11] See Wagstaff, A., (1989); and Rego G, N.R., (2010)

Examination of the data revealed that data for district hospitals (HOSP1) from Brazil constituted a significant share of the outpatient cost data for this facility type. We therefore decided to us a dummy for Brazil facilities level 3 (i.e., HOSP1).

Table 8 shows the final regression model for outpatient unit costs. As this represents a different approach to estimating the costs of outpatient visits than that used in the previous exercise, no comparison with the models of Adam et al. is possible.

Table 8. Regression output for outpatient unit cost expressed in 2007 Int. $

Dependent variable: Natural log of outpatient unit cost expressed in 2007 Int.$. (used to predict values for all levels, 1-5)

	Coefficient	95% confidence interval
Natural log of GDP per capita (PPP)	0.865***	[0.826,0.905]
Natural logarithm of outpatient visits	-0.0142*	[-0.0272,-0.00119]
Natural log of visits per provider per day (nurses , gps)	-0.0412***	[-0.0578,-0.0246]
Dummy variable for urban location of the facility	0.352***	[0.268,0.435]
Dummy variable for public level hospitals	-0.290***	[-0.330,-0.249]
Dummy variable for private level hospitals	0.0532*	[0.00479,0.102]
Dummy variable for level 2 facilities	0.208***	[0.144,0.271]
Dummy variable for level 3 facilities	0.304***	[0.213,0.395]
Dummy variable for level 4 facilities	0.348***	[0.279,0.417]
Dummy variable for observations in Colombia	0.628***	[0.542,0.713]
Dummy variable for observations in Brazil	-1.563***	[-1.656,-1.470]
Dummy variables for type 3 facilities in Brazil	-0.245***	[-0.337,-0.153]
Constant	-4.534***	[-4.797,-4.271]
Observations	9028	
R^2	0.658	
Adjusted R^2	0.658	
AIC	18820.9	
BIC	18913.3	
F_stat	1635.7	
Prob	0	
RESET	180.2	
RESET_P_value	1.75e-113	
Log_likelihood	-9397.4	
CF	1.271	
VIF	3.209	

*** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$**

4.3 Discussion

4.3.1 Imputation

For many reasons, missing data were a common feature of our data sets, and imputation is a generally used method which fills in the missing data using observed values according to various methods.

Tables 5A and 5B above in section 3 show the extent to which data was missing. Data on the number of visitors per provider per day, drug cost and bed occupation rate were the variables with the fewest available observations. Although we developed several models that applied a standard imputation function assuming the data were missing completely at random, these models performed less well according to statistical and validation tests than models not including imputed values. For this reason, the final regression models chosen did not use imputation. This probably indicates that data with missing values are less reliable

than other data, which further suggests that the 'missing completely at random' assumption is inappropriate.

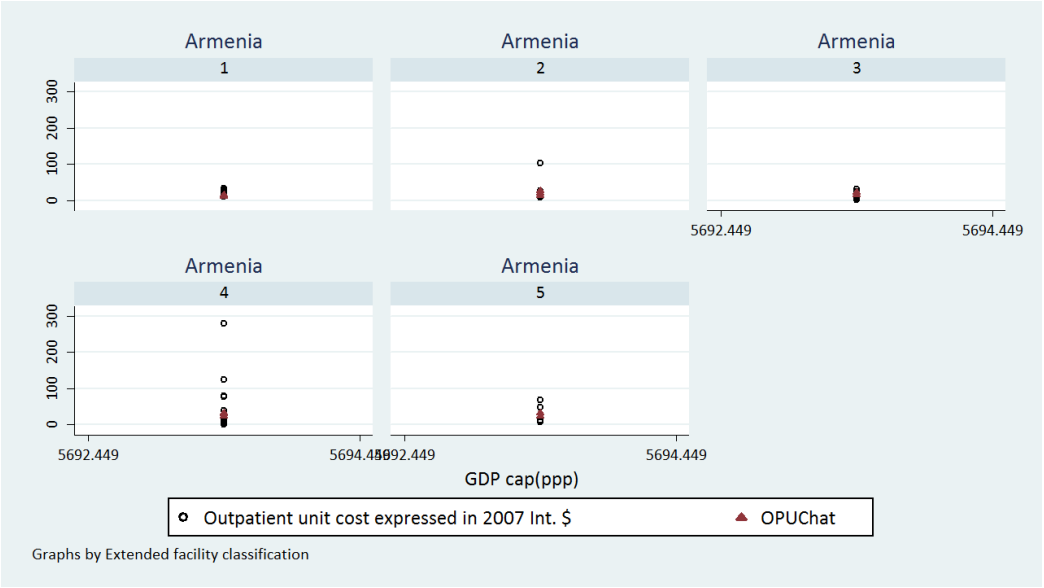
4.3.2 Model validation and goodness of fit

4.3.2.1 Tests employed

The tests used for judging model validity and the goodness of fit included the Breusch-Pagan/Cook-Weisberg test for heteroskedasticity, Ramsey's regression specification-error test for omitted variables, the tolerance test and its reciprocal variance inflation factor, plots of the residuals versus the fitted values, plots of the residuals versus the independent variables, plots of the predicted values versus the continuous independent variables, estimates of adjusted R-squared, the Akaike information criterion, the Bayesian information critierion, and F-statistics of the regression model.

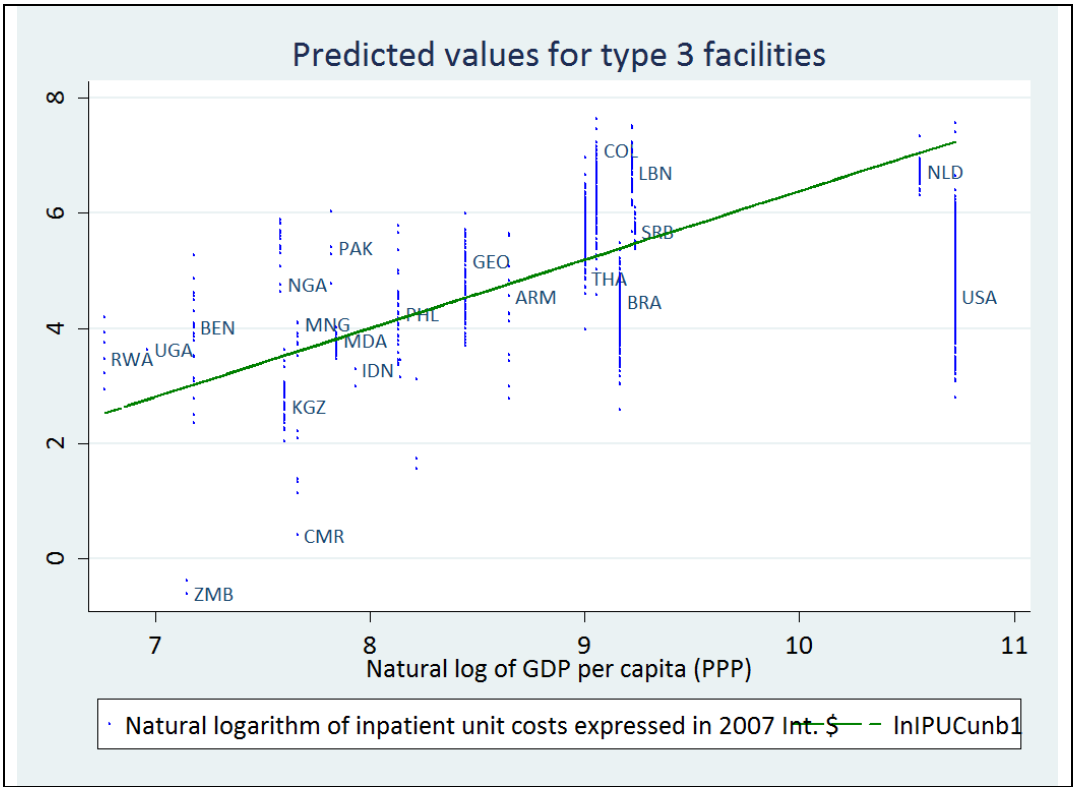
4.3.2.2 Internal validation tests

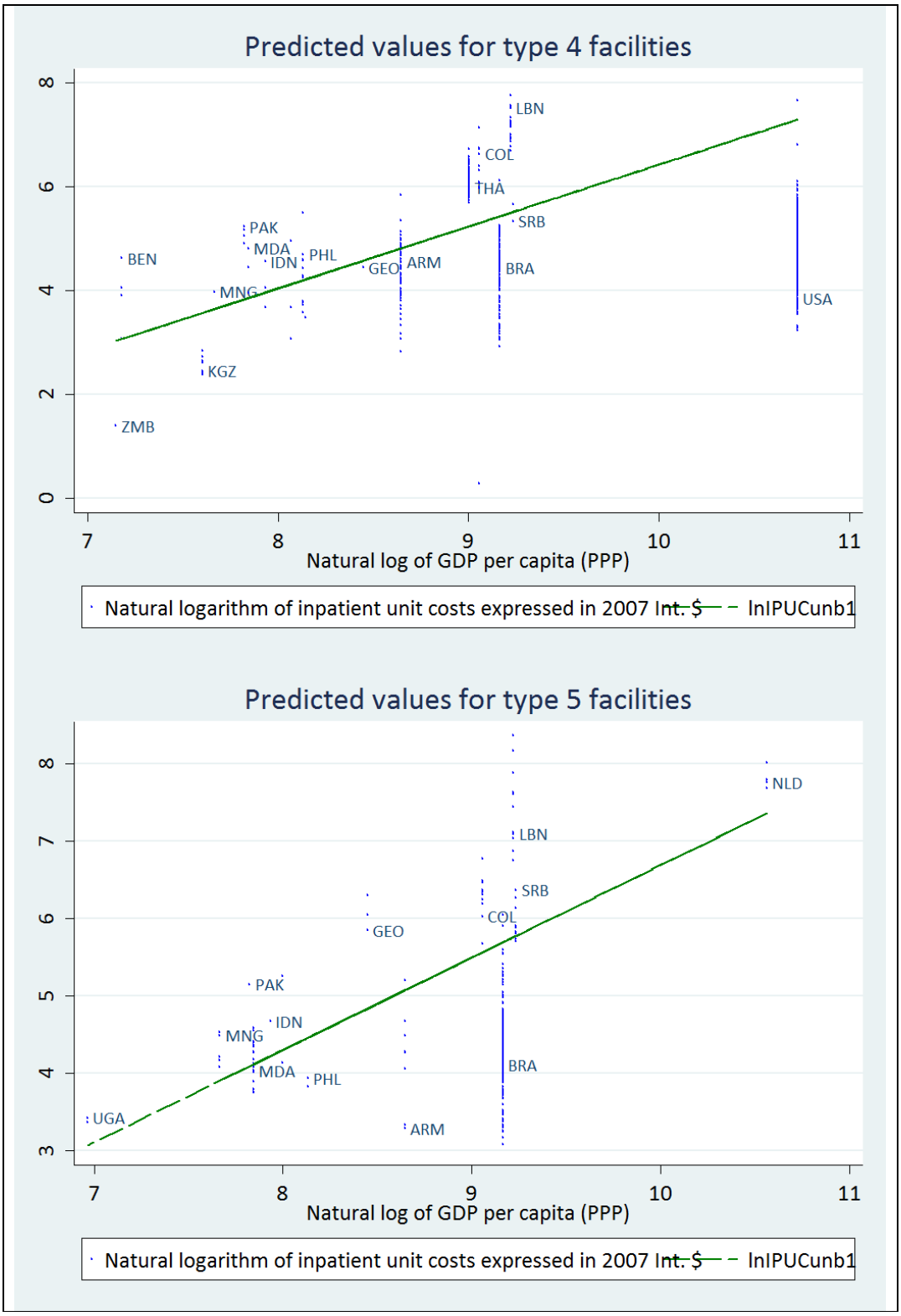
For internal validation purposes we used scatter graphs of actual and predicted observations by facility type for every country. Predicted values were based on the 80th percentile of variables used for prediction. Graphs for an example country (Armenia) are shown below.



4.3.2.3 Graphs of the predicted values versus GDP for inpatient unit cost

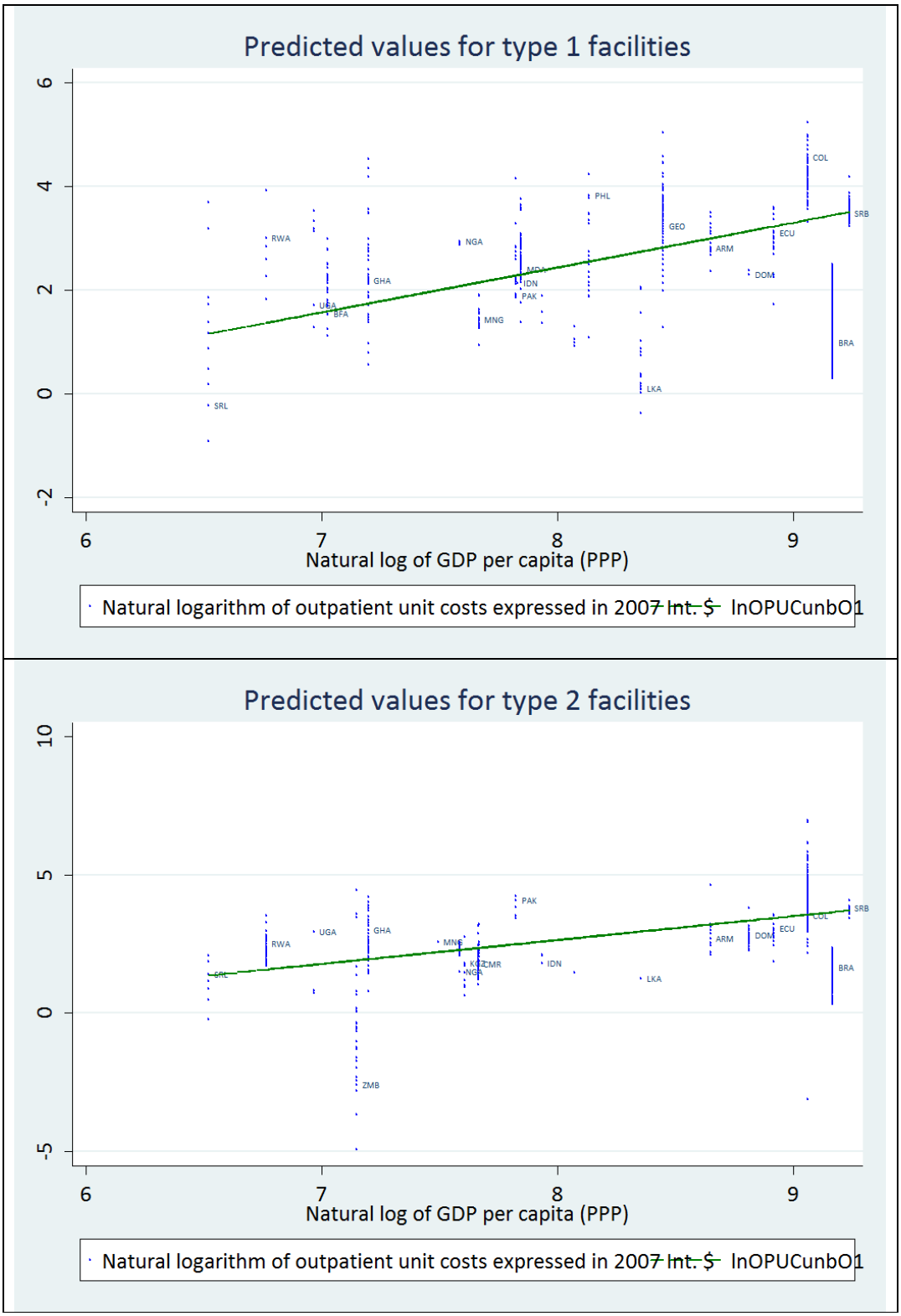
Graphs of the predicted values (using the 80th percentile value of predictor variables) of the unit cost per inpatient day versus GDP, stratified by facility type, are shown in the following figures.

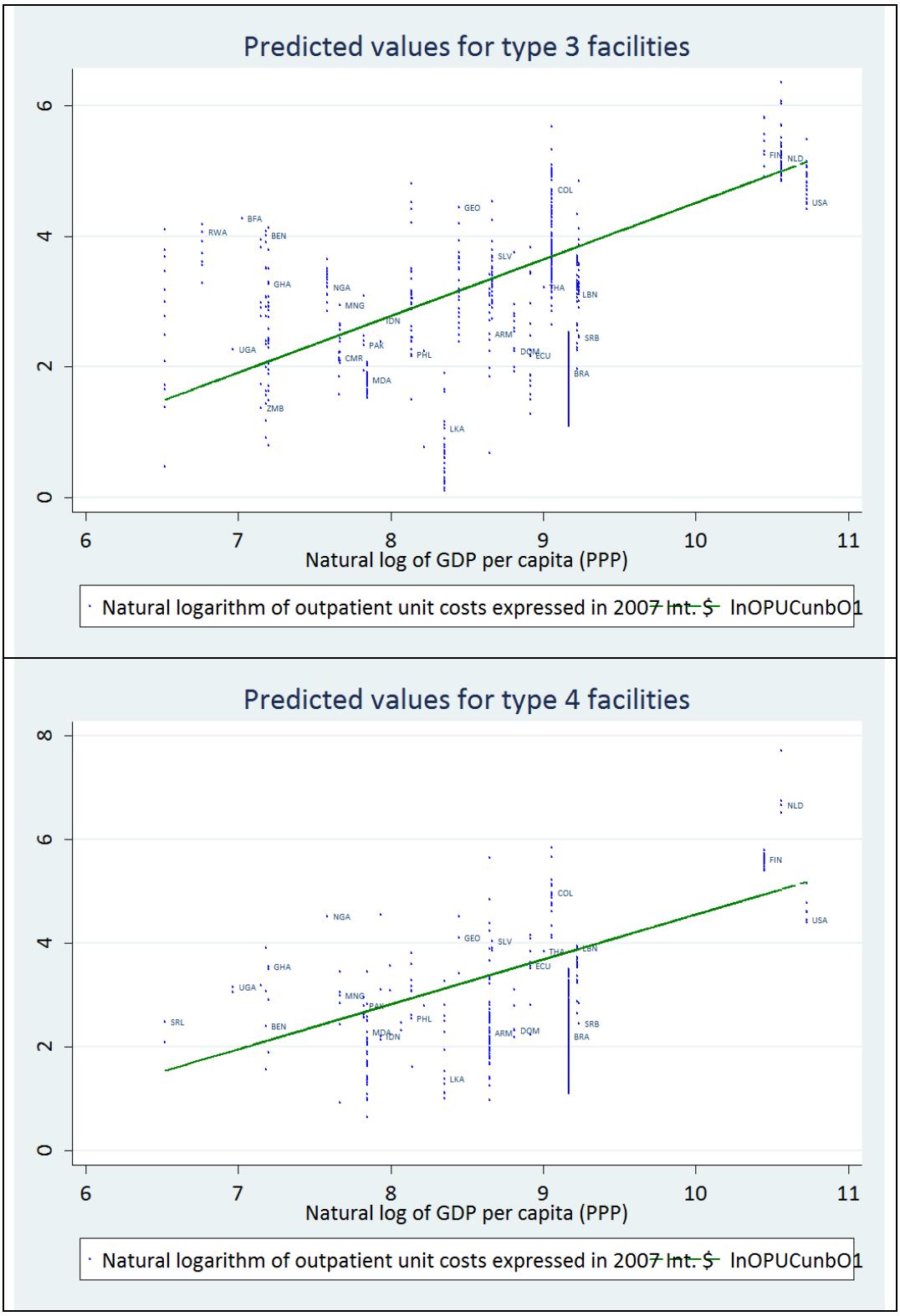




4.3.2.4 Graphs of the predicted values versus GDP for outpatient unit cost

Graphs of the predicted values (using the 80th percentile value of predictor variables) of the unit cost of outpatient visits versus GDP, stratified by facility type, are shown in the following figures.





4.4. Prediction of unit costs

The regression models can be used to predict inpatient day and outpatient visit costs in countries where direct observations are not available. Even in countries where data were

obtained, the predicted values in principle represent 'average values' that have desirable properties that may not always be possessed by the sample observations for that country (Adam et al., 2008).

The resulting predictions can be set to represent different facility types and levels by assigning the appropriate values to the variables in the regression equation. Some variables need to be set to country-specific values (e.g. GDP, country-specific dummy variables). Other variables (e.g. occupancy rate, average length of stay) can be set to normative values, sample medians, or to other values as appropriate.

4.4.1 Prediction of inpatient unit cost

Cost per inpatient bed day is estimated for public hospitals with 80% occupancy rate. For the variables 'total admissions' and 'average length of stay' (ALOS) we used the 80th percentile (p80) values from the sample of facilities used for the final regression models, as shown in Table 9.

Table 9. Values of variables used for prediction of the unit cost of an inpatient day

Total admissions, inpatient	p80
Facility level 3	4971
Facility level 4	14028
Average length of stay, inpatient	p80
Facility level 3	7.14
Facility level 4	9.75

4.4.2 Prediction of outpatient unit cost

For 'total outpatient visits' per facility per year we used the 80th percentile values within our sample, specific to each level, as shown in Table 10.

Table 10. Values of the variables used for prediction of the unit cost of an outpatient cost (80th percentile).

Facility type	Total outpatient visits per facility per year	Visits/provider per day
Facility level 1	67,656	8.96
Facility level 2	46,434	9.52
Facility level 3	93,739	3.22
Facility level 4	281,156	2.36

Outpatient visit costs at level 5 are assumed to be the same as those for level 4.

It can be observed that the visits per provider per day are reported to be fairly low for levels 3 and 4. One reason for this is likely to do with data quality since this parameter is calculated within the database and not reported directly by the user. There is thus a risk that the data collectors have reported the entire number of staff in the facility, even though not all of them deal with outpatient care (e.g. in a higher level hospital). Therefore, if a significant proportion of time is spent on inpatient care, the derived number of outpatient visits/provider/day can be expected to be low.

Secondly, other things being equal, the higher the level of facility the larger the proportion of complicated cases that is expected to be seen. As such cases require more time, therefore fewer visits per provider are expected to take place.

4.4.3 Comparison with previous (2000) predictions

Regression models based on previous work performed at WHO (Adam et al. 2003, Adam & Evans 2006, Adam et al. 2008) were used to compare predicted values with those resulting from the current work. For the purposes of comparison, all prediction models were set to predict for a base year of 2007 denominated in international dollars.

4.4.3.1 Comparison of predictions of inpatient unit cost

Predictions of inpatient unit cost for 162 countries for which information on the predictor variables was readily available were compared, and summary statistics are shown in Table 11. 'First-level hospital' refers to a district or similar hospital; 'second-level hospital' refers to a referral hospital or a specialist provincial hospital. 'Third-level hospital' refers to a second-level hospital that is also a teaching hospital. As, apart from income level, facility level is one of the most important determinants of cost, the values were derived by calculating average predicted cost by facility level (Table 11).

Table 11. Comparison of inpatient unit cost predictions, per day of hospitalization

Statistic/level	Predictions using Adam et al. models for prediction (2007 I$)	2007 predictions (2007 I$)
Overall		
Average	79	175
Standard deviation	119	251
Coefficient of variation	150%	143%
Maximum	1677	1939
Minimum	3.9	1.4
First-level hospital (facility level 3)		
Average	58	155
Standard deviation	84	219
Coefficient of variation	144%	141%
Maximum	941	1431
Minimum	3.9	1.4
Second-level hospital (facility level 4)		
Average	76	161
Standard deviation	110	227
Coefficient of variation	144%	141%
Maximum	1228	1488
Minimum	5.1	1.5
Third-level hospital (facility level five)		
Average	104	210
Standard deviation	150	296
Coefficient of variation	144%	141%
Maximum	1677	1939
Minimum	7	1.9

The current predicted costs of inpatient bed days are higher than those based on the previous Adam et al. model. However, the relative variability (as measured by the coefficient of variation) is similar to that of the previous predictions.

The means of the predictions of inpatient unit cost by income category and facility level are shown in Table 12.

Table 12. Comparison of inpatient unit cost predictions by type and income class

Income Classification	Facility type			
	3	4	5	Total
Low	8.339	8.675	11.31	9.440
Lower-Middle	44.11	45.89	59.81	49.94
Upper-Middle	126.9	132.0	172.0	143.7
High	493.3	513.2	668.8	558.4
Total	154.6	160.8	209.6	175.0

4.4.3.2 Comparison of outpatient unit cost predictions

Predictions of outpatient unit cost were compared and summary statistics are shown in Table 13. Average values by income class and facility type are presented in table 14.

Table 13. Comparison of outpatient unit cost predictions, per patient visit

Statistic/level	Predictions using Adam et al. 2006 and 2008 models (2007 I$) *	2007 predictions (2007 I$)
Overall		
Average	20.7	19.8
Standard deviation	46.8	20.2
Coefficient of variation	226%	102%
Maximum	832.2	122.7
Minimum	0.5	0.7
Health center (facility level 1)		
Average	2.3	15.3
Standard deviation	0.8	15.4
Coefficient of variation	35%	100%
Maximum	3.9	83.7
Minimum	0.5	0.8
Health center with beds (facility level 2)		
Average	2.3 [a]	18.9
Standard deviation	0.8 [a]	19.0
Coefficient of variation	35% [a]	100%
Maximum	3.9 [a]	103.3
Minimum	0.5 [a]	0.9
First-level hospital		

(facility level 3)		
Average	21.9	19.8
Standard deviation	35.9	19.9
Coefficient of variation	164%	101%
Maximum	397	108.2
Minimum	0.9	1.0
Second-level hospital (facility level 4)		
Average	31.1	22.5
Standard deviation	50.9	22.6
Coefficient of variation	164%	100%
Maximum	562.6	122.7
Minimum	1.2	1.1
Third-level hospital (facility level 5)		
Average	46	23
Standard deviation	75.3	22.6
Coefficient of variation	164%	100%
Maximum	832.2	122.7
Minimum	1.8	1.1

[a] The Adam et al. model does not distinguish between first and second level facilities (health centers and health centers with beds).
[b] The 2007 outpatient unit cost regression model does not distinguish between second- and third-level hospitals. In order to predict values for second level hospitals, a value of 1 for the dummy for teaching hospitals was used.
* The 2000 model has been updated with recent values to allow for comparison.

Table 14. Comparison of outpatient unit cost predictions by type and income class

World Bank Income Classification	Facility type					
	1	**2**	**3**	**4**	**5**	**Total**
Low	2.41	2.98	3.12	3.54	3.54	3.12
Lower-Middle	7.76	9.58	10.04	11.38	11.38	10.03
Upper-Middle	15.69	19.38	20.27	23.01	23.01	20.27
High	39.74	49.07	51.40	58.27	58.27	51.35
Total	15.31	18.90	19.80	22.45	22.45	19.78

The average predicted cost for an outpatient visit in the current model is much higher (i.e. by a factor of approximately 5-8) at the health-center level compared to the previous model. At first-level hospitals, however, the average predicted cost is similar to that based on previous (2000) work. At second and third level hospitals, the 2007 estimates are lower than previous (2000) estimates.

5. Further work undertaken March-June 2011

A draft version of this report was shared with costing experts and stakeholders and a technical briefing was held on 11 March 2011 with the objective to present the new estimates to the GFATM and to colleagues within WHO and UNAIDS who make use of unit cost estimates for costing, cost-effectiveness and expenditure tracking purposes.

As a direct follow-up to the meeting, the most recent unit cost predictions were shared with colleagues from GFATM, WHO, and UNAIDS for feedback. The predicted CHOICE unit cost estimates were compared with unit cost data collected from various projects specifically looking at TB and HIV-related health services. In general such data is scarce, and estimates could only be compared for 7 countries. In general there was no conclusive trend, as for some countries the specific country studies showed higher estimates than the CHOICE predictions, whereas for other countries the country study costs were lower.

Overall, as shown by Adam et al., the use of single unit estimates may be misleading as they may not be representative of the overall health system. Similarly the exact parameters and representativeness of the STB and Futures provided estimates is not known. The overall purpose of the CHOICE estimation model is to enable estimates to be produced for settings where no local data is available. As such it has been developed to provide the best fit according to global data, but it will not predict perfectly for all countries. It is maintained that whenever good quality country unit cost data is available from a representative sample, this should be used rather than using the CHOICE predictions.

The unit cost estimates were subsequently updated with WHO data for GDP per capita and WHO data for exchange rates to US$ and I$. The unit cost estimates are considered the final "2007" estimates and will be made public on the WHO-CHOICE website. The estimation method will be subject to continual revision and peer review and the estimates may change over time if new evidence becomes available to improve the prediction model.

6. Conclusions, limitations and recommendations for further work

6.1. Consistency/changes in cost estimates from 2000 to 2007

As shown above, the cost of inpatient bed days are significantly higher when based on the current (2007) model predictions than for the original (2000) models. The cost of outpatient visits are higher for levels 1-3, and lower for levels 4 and 5.

6.2. Limitations

6.2.1. Data availability

Although about 80 proposals were submitted to WHO, in the end only about 30 were accepted, taking into account regional and income level distribution as well as the level of disaggregation of data as indicated by country researchers. Thus, fewer data sets than expected matched the requirements for the analysis. In fact, the original budget for data purchase from country contractors was not fully employed due to difficulties in locating suitable data. In consequence, there seems to be a lack of available unit cost information that takes into account the kind of detailed disaggregation of cost category that was requested for this study.

With regards to publicly available data sets, although we reviewed a number of them we found that few provided information according to the structure required for analysis. Thus, the objective of secondary analysis on data collected for other purposes was only partially achieved.

Another limitation is the lack of data on costs in health centres from high income countries (i.e. doctors' surgeries/clinics are probably the closest analog).

6.2.2. Data quality

Efforts were made to follow up with data providers to ensure that the data provided corresponded to standard definitions and requested criteria. Sustained follow up was often required. Discussions were held with data providers on methods to enhance the quality and representativeness of the data when the data was incomplete and/or not conforming to the requested specification.

In several cases, weaknesses in the data were not communicated by providers but rather were only revealed through specific probing questions; this might suggest the possible underutilization and limited analysis of their data by providers/owners.

6.2.3. Capacity and resource limitations of country providers

Data providers reported limited capacity to undertake specific analyses, often due to change sin personnel at the ministry of health as well as general data management problems. This study highlights the need to strengthen skills to understand and make use of data at country level, in particular in low-income countries where resources are limited and the use of economic and financial data for evaluating current system performance could lead to considerable efficiency gains. [12]

An overall challenge is that many datasets were missing key variables. As mentioned above, the average data completion rate for the entire sample of facilities was 56%, with some contractors providing information on less than 50% of variables.

6.2.4. Reporting

An additional limitation that was not anticipated was the lack of a standardized reporting format for the qualitative report which data providers were requested to submit along with their data.

6.3 Recommendations for further work

The updated estimates of facility-specific unit cost constitute an important addition to the WHO-CHOICE database. The estimates and models can be used for desk-based analyses as well as for estimates of the costs of service-delivery in country for HIV, TB and malaria programmes (notably, ART, DOTS, MDR-TB, and malaria treatment). The models should also provide valid estimates of the unit costs of service delivery through the general (i.e. non-programme-specific) health care system. Since estimates of costs for drugs and equipment have been extracted, these estimates can be combined with programme-specific costs for TB and malaria treatment in order to arrive at programme- and intervention-specific unit costs for services. It should be kept in mind however that estimates reflect a regression analysis and as such represent an average relationship between cost determinants .

For the medium term it is recommended that the unit costs should be updated systematically at regular intervals of 5-7 years in order to adequately reflect changes in

[12] Several studies from low and middle income countries indicate that considerable savings could be made from increasing hospital efficiencies (WHO, 2010)

technology and cost structure. The results from this study indicate that capacity needs to be strengthened among data providers/owners in country, to allow for greater adoption of standardized methods as well as a better understanding of the information collected and its potential use to inform planning and resource allocation. In any future updates, a detailed, standardized template for the written report to be submitted by data owners should be provided in order to facilitate analysis of the data provided and a better understanding of the context in which it was collected. Additional efforts should be made in the next round to specify that data is specifically needed on the drugs, lab and food categories of costs.

More research may be needed regarding the assumptions used for continuous independent predictor variables in order to derive country predicted values.

Over the medium to long term, it would be possible to envisage developing partnerships with academic institutions and collaborating centres, with WHO in the role of a knowledge partner, in order to build the local knowledge base and capacity in order to address some of the shortcomings identified above. This will help to expand the dataset to include more countries.

References

Adam, T., D.B. Evans, and C.J. Murray, Econometric estimation of country-specific hospital costs. Cost Eff Resour Alloc, 2003. 1(1): p. 3.

Adam, T. and D. Evans, Determinants of variation in the cost of inpatient... [Soc Sci Med. 2006] - PubMed result. Social Science & Medicine, 2006. 63: p. 1700-1710.

Adam T. et al., Capacity utilization and the cost of primary care visits: implications for the costs of scaling up health interventions. Cost Eff Resour Alloc, 2008. 6: p. 22.

Disease Control Priorities in Developing Countries (2nd Edition), (2006)

Duan N., Smearing estimate -- a nonparametric retransformation method. 1983: Journal of the American Statistical Association. p. p. 605-610.

Hutubessy RC, Baltussen RM. Tan-Torres Edejer T, and Evans DB. 2003."Generalised Cost-Effectiveness Analysis: An Aid to Decision Making in Health." In Making Choices in Health: WHO *Guide to Cost-Effectiveness Analysis*, ed. T. Tan-Torres Edejer, R. M. Baltussen, T. Adam, R. Hutubessy, A. Acharya, D. B. Evan, and C. J. L. Murray, 277-88. Geneva: World Health Organization.

Pauly MV. "Estimating hospital costs." Journal of Health Economics, Vol. 5: 107-127, 1986.

Rego G, Costa JNR. The challenge of corporatisation: the experience of Portuguese public hospitals. 2010: Eur J Health Econ. p. 367-81.

Stanciole AE, Tan-Torres Edejer T, and Georgios Gkountouras. Unit Costs of Health Care Services of General Utilization: A Review of the International Literature and Current Estimates, 2009.

Tsilaajav, T., Costing study for selected hospitals in the Philippines, 2009.

Wagstaff A. Econometric studies in health economics : A survey of the British literature. Journal of Health Economics, 1989. 8.

WHO, World Health Report 2010: Health systems financing: the path to universal coverage.

Annex 1. Scoping questionnaire of available datasets

WORLD HEALTH ORGANIZATION
DEPARTMENT OF HEALTH SYSTEMS FINANCING

Unit Costs of Health Facilities 2009 Study: Scoping of available datasets

Name:_____
Institutional Affiliation:_____
E-mail address:_____
Telephone number:_____
Country that is the source of data_____

- When was the unit cost data collected? _____
  (anytime from 2000 to the present)

- How many health care facilities for which unit cost data are available:
  a. Health facilities only with outpatient services ____
  b. Health facilities primarily for outpatient services but with a limited number of day beds _____
  c. Health facilities primarily for inpatient services for simple cases (district hospitals, etc) _____
  d. Health facilities primarily for inpatient services for referral cases _____

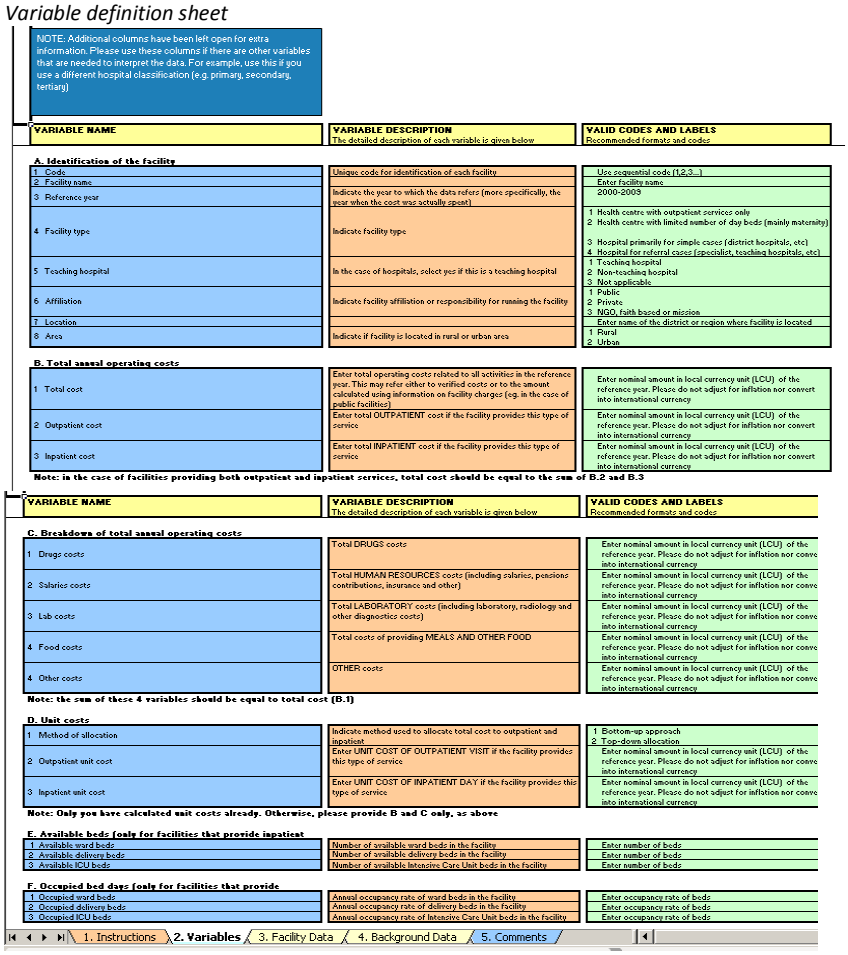
- Could you provide data, for each facility surveyed, on:
  a. number of bed days Y/N
  b. number of admissions Y/N
  c. number of outpatient visits Y/N
  d. number of ancillary services Y/N

- Could you provide data on determinants of costs and efficiency, for each facility surveyed, particularly
  a. number of beds Y/N
  b. average length of stay Y/N
  c. occupancy rate Y/N
  d. number of medical staff per bed Y/N
  e. number of outpatient visits per medical staff per day Y/N

  Please provide a list of additional variables which are available per facility in the dataset (e.g. number of specialist physicians in the facility, number of surgical procedures performed, etc). Or attach list of variables in a separate page.
  _____
  _____
  _____
  _____
  _____

Annex 2. Survey questionnaire

Variable definition sheet



NOTE. Additional columns have been left open for extra information. Please use these columns if there are other variables that are needed to interpret the data. For example, use this if you use a different hospital classification (e.g. primary, secondary, tertiary)

VARIABLE NAME	VARIABLE DESCRIPTION The detailed description of each variable is given below	VALID CODES AND LABELS Recommended formats and codes
A. Identification of the facility		
1 Code	Unique code for identification of each facility	Use sequential code (1,2,3...)
2 Facility name		Enter facility name
3 Reference year	Indicate the year to which the data refers (more specifically, the year when the cost was actually spent)	2000-2003
4 Facility type	Indicate facility type	1 Health centre with outpatient services only 2 Health centre with limited number of day beds (mainly maternity) 3 Hospital primarily for simple cases (district hospitals, etc) 4 Hospital for referral cases (specialist, teaching hospitals, etc)
5 Teaching hospital	In the case of hospitals, select yes if this is a teaching hospital	1 Teaching hospital 2 Non-teaching hospital 3 Not applicable
6 Affiliation	Indicate facility affiliation or responsibility for running the facility	1 Public 2 Private 3 NGO, faith based or mission
7 Location		Enter name of the district or region where facility is located
8 Area	Indicate if facility is located in rural or urban area	1 Rural 2 Urban
B. Total annual operating costs		
1 Total cost	Enter total operating costs related to all activities in the reference year. This may refer either to verified costs or to the amount calculated using information on facility charges (eg. in the case of public facilities)	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
2 Outpatient cost	Enter total OUTPATIENT cost if the facility provides this type of service	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
3 Inpatient cost	Enter total INPATIENT cost if the facility provides this type of service	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency

Note: in the case of facilities providing both outpatient and inpatient services, total cost should be equal to the sum of B.2 and B.3

VARIABLE NAME	VARIABLE DESCRIPTION The detailed description of each variable is given below	VALID CODES AND LABELS Recommended formats and codes
C. Breakdown of total annual operating costs		
1 Drugs costs	Total DRUGS costs	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
2 Salaries costs	Total HUMAN RESOURCES costs (including salaries, pensions contributions, insurance and other)	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
3 Lab costs	Total LABORATORY costs (including laboratory, radiology and other diagnostics costs)	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
4 Food costs	Total costs of providing MEALS AND OTHER FOOD	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
4 Other costs	OTHER costs	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency

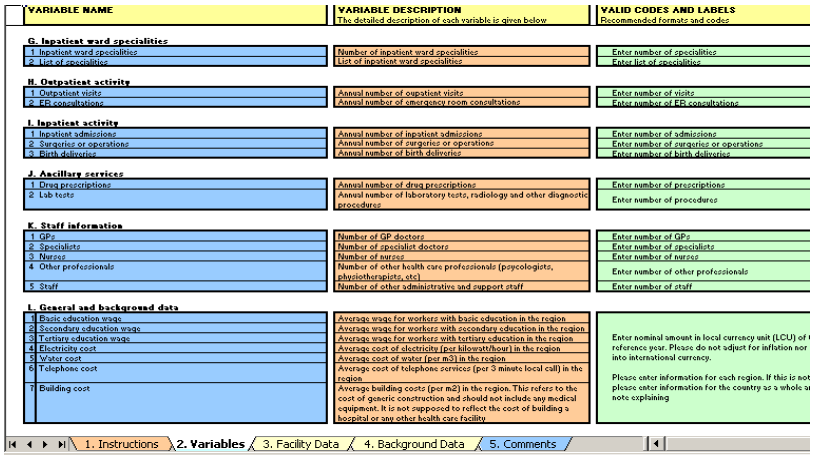
Note: the sum of these 4 variables should be equal to total cost (B.1)

D. Unit costs		
1 Method of allocation	Indicate method used to allocate total cost to outpatient and inpatient	1 Bottom-up approach 2 Top-down allocation
2 Outpatient unit cost	Enter UNIT COST OF OUTPATIENT VISIT if the facility provides this type of service	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency
3 Inpatient unit cost	Enter UNIT COST OF INPATIENT DAY if the facility provides this type of service	Enter nominal amount in local currency unit (LCU) of the reference year. Please do not adjust for inflation nor convert into international currency

Note: Only you have calculated unit costs already. Otherwise, please provide B and C only, as above

E. Available beds (only for facilities that provide inpatient		
1 Available ward beds	Number of available ward beds in the facility	Enter number of beds
2 Available delivery beds	Number of available delivery beds in the facility	Enter number of beds
3 Available ICU beds	Number of available Intensive Care Unit beds in the facility	Enter number of beds
F. Occupied bed days (only for facilities that provide		
1 Occupied ward beds	Annual occupancy rate of ward beds in the facility	Enter occupancy rate of beds
2 Occupied delivery beds	Annual occupancy rate of delivery beds in the facility	Enter occupancy rate of beds
3 Occupied ICU beds	Annual occupancy rate of Intensive Care Unit beds in the facility	Enter occupancy rate of beds

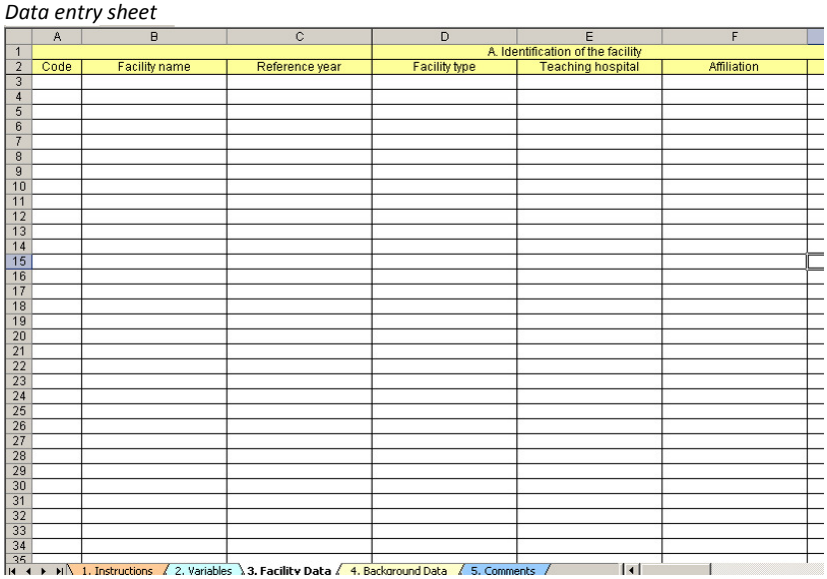
1. Instructions | 2. Variables | 3. Facility Data | 4. Background Data | 5. Comments



VARIABLE NAME	VARIABLE DESCRIPTION The detailed description of each variable is given below	VALID CODES AND LABELS Recommended formats and codes
G. Inpatient ward specialities		
1 Inpatient ward specialities	Number of inpatient ward specialities	Enter number of specialities
2 List of specialities	List of inpatient ward specialities	Enter list of specialities
H. Outpatient activity		
1 Outpatient visits	Annual number of outpatient visits	Enter number of visits
2 ER consultations	Annual number of emergency room consultations	Enter number of ER consultations
I. Inpatient activity		
1 Inpatient admissions	Annual number of inpatient admissions	Enter number of admissions
2 Surgeries or operations	Annual number of surgeries or operations	Enter number of surgeries or operations
3 Birth deliveries	Annual number of birth deliveries	Enter number of birth deliveries
J. Ancillary services		
1 Drug prescriptions	Annual number of drug prescriptions	Enter number of prescriptions
2 Lab tests	Annual number of laboratory tests, radiology and other diagnostic procedures	Enter number of procedures
K. Staff information		
1 GPs	Number of GP doctors	Enter number of GPs
2 Specialists	Number of specialist doctors	Enter number of specialists
3 Nurses	Number of nurses	Enter number of nurses
4 Other professionals	Number of other health care professionals (psycologists, physiotherapists, etc)	Enter number of other professionals
5 Staff	Number of other administrative and support staff	Enter number of staff
L. General and background data		
1 Basic education wage	Average wage for workers with basic education in the region	Enter nominal amount in local currency unit (LCU) of t reference year. Please do not adjust for inflation nor into international currency. Please enter information for each region. If this is not please enter information for the country as a whole a note explaining
2 Secondary education wage	Average wage for workers with secondary education in the region	
3 Tertiary education wage	Average wage for workers with tertiary education in the region	
4 Electricity cost	Average cost of electricity (per kilowatt/hour) in the region	
5 Water cost	Average cost of water (per m3) in the region	
6 Telephone cost	Average cost of telephone services (per 3 minute local call) in the region	
7 Building cost	Average building costs (per m2) in the region. This refers to the cost of generic construction and should not include any medical equipment. It is not supposed to reflect the cost of building a hospital or any other health care facility	

1. Instructions | 2. Variables | 3. Facility Data | 4. Background Data | 5. Comments

Data entry sheet



	A	B	C	D	E	F
1				A. Identification of the facility		
2	Code	Facility name	Reference year	Facility type	Teaching hospital	Affiliation
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						

1. Instructions | 2. Variables | 3. Facility Data | 4. Background Data | 5. Comments

Annex 3. Data points by reference year, country and facility type

Reference year	Country	Type 1	Type 2	Type 3	Type 4	Grand total
2000	El Salvador	3	27			30
	Mongolia			1		1
	Thailand		2			2
Total - 2000		3	29	1		33
2001	El Salvador	3	27			30
	Indonesia			1	4	5
	Thailand	3			55	58
Total - 2001		6	27	1	59	93
2002	Cameroon		6	39		45
	El Salvador	3	27			30
	Finland		1			1
	Thailand	7	12		11	30
Total - 2002		10	46	39	11	106
2003	El Salvador	3	27			30
	Indonesia	4	2	4	3	13
	Mongolia	6	10	8	14	38
	Thailand		735		42	777
Total - 2003		13	774	12	59	858
2004	Dominican Republic	1	4	8	1	14
	El Salvador	3	27			30
	Indonesia	2				2
Total - 2004		6	31	8	1	46
2005	Burkina Faso		1		24	25
	Dominican Republic	2	6	18		26
	El Salvador	3	27			30
	Indonesia	3	1	1	4	9
	Sri Lanka	16	36	1	16	69
Total - 2005		24	71	20	44	159
2006	Burkina Faso				1	1
	Dominican Republic	2	5	13	1	21
	El Salvador	3	27			30
	Indonesia	1	3			4
	Rwanda		7	30	6	43
	Thailand		1			1
	Zambia	2	15	124	3	144
Total - 2006		8	58	167	11	244
2007	Armenia	90	16	16	16	138
	Australia	21	55			76
	Benin	5	24			29
	Brazil	482	1680	206	6519	8887
	Colombia	26	128	688	83	925
	Ecuador	8	18	18	18	62
	El Salvador	6	54			60
	Finland	20	16			36
	Georgia	4	97	14	233	348
	Ghana	5	36	90	68	199
	Indonesia	1	5			6
	Kyrgyzstan	7	47	9		63

	Lebanon	**33**	**48**			**81**
	Netherlands	**5**	**32**			**37**
	Nigeria	**1**	**16**	**8**	**5**	**30**
	Pakistan	**5**	**5**	**5**	**20**	**35**
	Philippines	**14**	**31**	**2**	**21**	**68**
	Republic of Moldova	**29**	**34**		**84**	**147**
	Serbia	**23**	**23**	**10**	**50**	**106**
	Sierra Leone	**8**	**34**	**175**	**781**	**998**
	Thailand	**93**	**4**		**91**	**188**
	Uganda	**2**	**1**	**3**	**6**	**12**
	United States of America	**1346**	**3592**			**4938**
Total - 2007		**2234**	**5996**	**1244**	**7995**	**17469**
Grand total		**2304**	**7032**	**1492**	**8180**	**19008**

Annex 4. Regression models used in past WHO-CHOICE Unit cost analysis

Adam & Evans model for inpatient unit costs (2003)

	Coefficient
Natural log of GDP per capita (PPP)	0.7624
Natural log of occupancy rate	-0.2318
Drug costs	0.641
Food costs	0.2116
Dummy variable for private hospitals	0.2444
Dummy variable for public hospitals	-0.2722
Dummy variable for level 1 hospital *	-0.5777
Dummy variable for level 2 hospital **	-0.3118
Dummy variable for observations of United States	1.7471
Constant	-2.5036
Correction factor	1.25

*corresponds to level 3 in our updated analysis
** corresponds to level 4 in our updated analysis

Adam & Evans model for outpatient and inpatient unit costs (2006)

	Coefficient
Ln GDP per capita	**0.1303**
Ln occupancy rate	**0.1683**
Ln hospital beds	**0.0884**
Public	**-0.489**
Food costs	**-0.1985**
Sri Lanka	**-1.2401**
Thailand	**-0.2998**
China	**0.3449**
Ecuador	**-0.8187**
Constant	**-2.2698**
Correction factor	**1.13**

Adam et al. model for outpatient unit costs (2008)

	Coefficient
Natural log of GDP per capita (PPP)	**0.6219**
Natural log of visits per provider per day (nurses, GPs, specialists)	**-0.2756**
Capital costs	**0.7759**
Dummy variable for former communist regime	**-0.466**
Dummy variable for public hospitals	**-0.2541**
Constant	**-2.906**
Correction factor	**1.45**

Annex 5. Restrictions imposed ('outlier' elimination)

Outliers were detected with the help of histograms and scatter diagrams of explanatory and dependent variables versus GDP/capita. Boundaries for acceptable values were identified either by graph inspection (i.e removing observations that were observed to be located in the extremes of a distribution) or based on logical a-priori limits regarding the possible magnitude of values for the explanatory and dependent variables . Different restrictions were imposed for the IPUC and the OPUC models. For each restriction we observed the effect on the number of observations remaining within the model. In general the restrictions imposed had no significant impact on the number of observations included in the final regression models (again, suggesting that observations with missing data were more unreliable than others). The following table shows the restrictions that were imposed and their impact.

IPUC model		Restriction	Result on total observations	Observation left for the model	Impact on the model observations
		No restriction	0 observations deleted	3653	
	IP1	lninpatient>=9 & type==4	232 observations deleted	3653	0
	IP2	lninpatient>=8 & type==3	1167 observations deleted	3653	0
	IP3	Lnalos>=3	9779 observations deleted	3570	-83
	IP4	Lnalos<=-2	2 observations deleted	3570	0
	IP5	Lnpctwardbeds>2	131 observations deleted	3570	0
	IP6	Lnpctwardbeds<-5	83 observations deleted	3499	-71
	IP7	lnadmissions<=4.60517	223 observations deleted	3479	-20
	IP8	totalbeds>=4000	73 observations deleted	3410	-69
	IP9	country=="Zambia"	3 observations deleted	3407	-3
OPUC model					
		No restriction	0 observations deleted	9234	
	OP1	type==4 & lnoutpatient<=0	3 observations deleted	9231	-3
	OP2	type==3 & lnoutpatient<=0	4 observations deleted	9227	-4
	OP3	type==2 & lnoutpatient>=8	102 observations deleted	9226	-1
	OP4	type==1 & lnoutpatient>=7	11 observations deleted	9225	-1
	OP5	lnviz2<=-5	100 observations deleted	9125	-100
	OP6	viz2>200	9664 observations deleted	9107	-18
	OP7	lnvisits<=0	0 observations deleted	9107	0
	OP8	totalstaff2>=7000	4 observations deleted	9103	-4
	OP9	visits>=10^6	75 observations deleted	9028	-75