Limitations of methods for measuring out-of-pocket and catastrophic private health expenditures

Chunling Lu,[a] Brian Chin,[b] Guohong Li[c] & Christopher JL Murray[d]

Objective To investigate the effect of survey design, specifically the number of items and recall period, on estimates of household out-of-pocket and catastrophic expenditure on health.

Methods We used results from two surveys – the World Health Survey and the Living Standards Measurement Study – that asked the same respondents about health expenditures in different ways. Data from the World Health Survey were used to compare estimates of average annual out-of-pocket spending on health care derived from a single-item and from an eight-item measure. This was done by calculating the ratio of the average obtained with the single-item measure to that obtained with the eight-item measure. Estimates of catastrophic spending from the two measures were also compared. Data from the Living Standards Measurement Study from three countries (Bulgaria, Jamaica and Nepal) with different recall periods and varying numbers of items in different modules were used to compare estimates of average annual out-of-pocket spending derived using various methods.

Findings In most countries, a lower level of disaggregation (i.e. fewer items) gave a lower estimate for average health spending, and a shorter recall period yielded a larger estimate. However, when the effects of aggregation and recall period are combined, it is difficult to predict which of the two has the greater influence.

Conclusion The magnitude of both out-of-pocket and catastrophic spending on health is affected by the choice of recall period and the number of items. Thus, it is crucial to establish a method to generate valid, reliable and comparable information on private health spending.

Une traduction en français de ce résumé figure à la fin de l'article. Al final del artículo se facilita una traducción al español. الترجمة العربية لهذه الخلاصة في نهاية النص الكامل لهذه المقالة.

Introduction

Valid, reliable and comparable information on national and international resource inputs for health is critical for developing health policies, managing programme implementation and evaluating efficiency and performance. Out-of-pocket payments incurred by households for medical services received (excluding transportation spending and insurance payments and reimbursements) are estimated to account for 23% of total global health expenditure and 45% of health expenditure in the developing world. Within the latter, out-of-pocket health spending ranged from 1.6% of total health expenditure in Niue to 82.9% in Guinea in 2003.[1] Over the past 6 years analysts have also suggested that out-of-pocket health spending is catastrophic for many households, often pushing them below the poverty line.[2–8] A household's health expenditure is considered to be catastrophic if the ratio between the household's out-of-pocket health expenditure and its disposable income reaches a certain critical point; commonly used thresholds include 30% or 40% of capacity to pay, or 10% of total expenditures.[3–5] The problem of catastrophic and impoverishing health payments has captured policy attention[9,10] and has led to major legislation[11] and system reform.[12–18]

The capacity to monitor and track meaningful change in out-of-pocket health spending and catastrophic payments is very limited. However, household surveys that include questions on different types of health expenditures and total health expenditure can help meet these information needs. Regular income and expenditure surveys – already widely used to support computation for national accounts[19,20] – collect this information, as do some international survey programmes, including The World Bank's Living Standards Measurement Study and WHO's World Health Survey. Unfortunately, these surveys vary in the exact wording of questions, the number of disaggregated expenditure categories, the recall periods and the framing of the expenditure questions; there is also variation within the same survey across different countries and years. The validity, reliability and comparability of information on out-of-pocket health spending gathered through such disparate methods have not been established.

Studies on the reliability and validity of total expenditure data have highlighted at least two factors that influence the results: the number of expenditure categories used and the recall period.[21–27] Even though the results are sensitive to the level of disaggregation, the number of items collected in published consumption surveys ranges from 1 to 1300.[28] Few validation studies have been undertaken in developing countries, and no studies have explored the issue of how to collect valid, reliable and comparable information on health expenditures.

In this paper we use the World Health Survey and the Living Standards Measurement Study – two household surveys which asked the same respondents about health expenditures in different ways – to explore two sources of potential bias: the

[a] Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States of America (USA).
[b] Population Studies Center, University of Pennsylvania, Philadelphia, PA, USA.
[c] School of Public Health, Shanghai Jiao Tong University, Shanghai, China.
[d] Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA.
Correspondence to Chunling Lu (e-mail: chunling_lu@hms.harvard.edu).

number of health expenditure categories and the recall period. Based on our findings, we discuss potential solutions to the problem of comparability.

Methods

World Health Survey

The World Health Survey was conducted by WHO using a consistent survey instrument in 50 developing countries between 2002 and 2004.[29] This survey first asks a single question on household health spending in the previous 4 weeks, so that recall of the disaggregated categories does not influence the response. Eight more detailed questions follow, focused on health spending in the same period. These questions elicit information on payments for outpatient services, hospitalization, traditional medicine services, dentists, medication, medical tests, health-care products and other expenditures. The health spending estimates can be derived from either the single-item or eight-item questions. Another question concerns inpatient costs in the previous 11 months (excluding the most recent month). Table 1 gives details of the health expenditure items. Countries that did not include all of the items listed in Table 1 in the survey (Hungary and Turkey) or that were missing more than 90% of the data on these items (Guatemala) were excluded from the analysis. Countries where 75% or more households reported the same amount of positive health spending with the single-item and eight-item measures (Brazil, Kazakhstan, Mauritius and Paraguay) were excluded from the analysis on the assumption that a high percentage of exact agreement between the two measures indicated a serious problem with data quality. Our final analysis thus included 43 countries. Table 2 (available at: http://www.who.int/bulletin/volumes/87/03/08-054379/en/index.html) shows the number of households surveyed and the response rates for these countries.

Since the level of disaggregation (number of questions) is the only difference between estimates of household health spending based on the single-item responses and on the eight-item responses, it is possible to detect the effect of disaggregation on the estimates of health spending for each country. We compared the average annual health spending estimate obtained from the single-item measure with that obtained

Table 1. **Number of items on health spending in the World Health Survey**

	Items
Recall period: previous 4 weeks	
Total out-of-pocket health spending	1
Out-of-pocket health spending	8
Hospitalization	1
Outpatient	1
Traditional medicine	1
Dentists	1
Medication	1
Health-care product	1
Laboratory test	1
Other	1
Recall period: previous 11 months excluding the last month	
Hospitalization	1

from the eight-item measure by calculating the ratio of the two averages. This ratio indicates which measure generates a larger estimate. We estimated the 95% confidence interval (CI) of the ratio using "bootstrapping" – a technique for generating a description of the sampling properties of empirical estimators using the sample data. To do this, we constructed several re-samples of the observed data set (of equal size to the observed data set), each of which was obtained by random sampling with replacement from the original data set. We also compared the estimates of catastrophic spending from the two measures, and examined how the level of disaggregation may affect the estimates. In this study, a household's expenditure was defined as catastrophic if the ratio between the household's out-of-pocket health expenditure and its capacity to pay, defined as effective income remaining after subsistence needs had been met, reached 0.4.[3,4] The numbers of observations used in the calculation are listed in Table 2.

Living Standards Measurement Study

The Living Standards Measurement Study, conducted by The World Bank,[30] is an important tool for measuring poverty in developing countries. It includes questionnaires designed to study various aspects of household consumption behaviour, including spending on health care. The study collects information on health spending in all selected countries, but the way the information is collected varies substantially. For example, some Living Standards Measurement

Study surveys collect health spending information at the household level using a consumption module, or at the individual level using a health module. Others collect this information only at the household level. While the Living Standards Measurement Study generally asks about household health spending over a 12-month recall period, for individual-level health spending the recall period varies from 2 weeks (Ghana 1999) to 12 months (India 1998). The number and specificity of the categories of health spending for which the study collects information also vary considerably. For example, the Living Standards Measurement Study surveys for China 1997, Guatemala 2000, India 1998 and the United Republic of Tanzania 2004 asked for an aggregate estimate of total health expenditures by household, whereas the survey for Bulgaria 2001 asked for household-level expenditure across six categories of health spending (outpatient care, inpatient care, dental care, medicines, optical equipment, and skin care and plastic surgery). These inconsistencies in the level of disaggregation are also present across the individual-level health spending modules. Such variations in how health spending information is collected enabled us to detect the sensitivity of household health spending estimates to different recall periods and levels of disaggregation.

We selected three Living Standards Measurement Study surveys with questions on health spending in both their consumption and health modules: Bulgaria 2001, Jamaica 2001 and Nepal 1997. Details on the recall period, number of items and type of modules are

presented for each country in Table 3, Table 4 and Table 5, respectively. The tables show that the surveys in the three countries varied in the number of items used to collect information in the health and consumption modules, and in the timing of the recall period. We used *t*-tests to compare the average annual health spending from the different recall periods and items and to examine the effects of these factors.

Results

Effect of disaggregation

Fig. 1 (available at: http://www.who.int/bulletin/volumes/87/03/08-054379/en/index.html) compares the ratio of the estimated average annual out-of-pocket health expenditure obtained from the single-item measure to that obtained from the eight-item measure in the World Health Survey. The ratio varied from 0.25 to 1.37. Among the 43 countries studied, 38 had ratios less than 1, with the difference from 1 being significant at the 0.95 confidence level in 37 of these cases. Thus, the single-item measure yielded a significantly lower estimate than the eight-item measure (the ratio in the Congo was not significantly different from 1). Among the remaining five countries (Mali, Namibia, Nepal, Sri Lanka and Uruguay), the ratio was significantly greater than 1. Thus, in most countries, a lower level of disaggregation gave a lower estimate for average health spending, a finding consistent with the results of previous studies on total household expenditure in developed and developing countries.[21,31,32] However, this finding

is not universally true across countries, and the degree of bias in the single-item method is highly variable.

Effect of recall period

The World Health Survey includes two questions on the costs of hospitalization, one with a 4-week recall period and the other with an 11-month recall period. Fig. 2 (available at: http://www.who.int/bulletin/volumes/87/03/08-054379/en/index.html) presents the ratio of the average annual household out-of-pocket spending on hospitalization derived from a 1-month recall period to that derived from an 11-month recall period. Among the 43 countries, 39 had ratios significantly greater than 1 at the 0.95

confidence level, with the highest ratio being 9.56. Four countries had ratios significantly less than 1. Thus, in most countries, a shorter recall period yielded larger estimates for average annual health spending. The variation as a function of recall period is enormous and raises serious doubts about the comparability of results from surveys that use different recall periods.

The Nepal Living Standards Measurement Study 1997 asked two questions on health spending twice in the consumption module, using first a 1-month and then a 12-month recall period. With a sample size of 2421, the means were 2490 Nepalese rupees (NRs) (95% CI: 2017–2962) from a 1-month recall and NRs 1887 (95% CI: 1697–2076) from a 12-month recall, a difference that is significant at the 0.95 confidence level. Thus, in Nepal, a short recall period appeared to result in a significantly larger estimate of the household health spending than a long recall period.

Effect of combined factors

In the Jamaica Living Standards Measurement Study 2001, with a sample size of 1665, the mean yearly out-of-pocket health spending at the household level was about 8944 Jamaican dollars (J$) (95% CI: 7433–10 455) from a health module with six items and a 4-week recall period. This figure was J$ 7174 (95% CI: 6270–8079) when derived from a consumption module with two

Table 3. Recall period and number of questions, Bulgaria Living Standards Measurement Study, 2001

	Recall period	
	Per month	12 months
Health module (individual)		
Costs for disability, chronic disease or ailment	1	
Consultation and test for injury, disease or ailment	1	
Treatment and drug for injury, disease or ailment	1	
Transportation	1	
Extra payments	1	
Consumption module (household)		
Dentist		1
Doctor		1
Hospital		1
Medicines		1
Optical equipment		1
Skin care/plastic surgery		1
Other health/specify		1

Table 4. Recall period and number of questions, Jamaica Living Standards Measurement Study, 2001

	Recall period	
	Past 4 weeks	12 months
Health module (individual)		
Costs in public health facility	1	
Costs in private health facility	1	
Costs in public hospital	1	
Costs in private hospital	1	
Medicine in public sources	1	
Medicine in private sources	1	
Consumption module (household)		
Medicine		1
Medical services (physician fees, hospital care, prescriptions, spectacles, etc.)		1

Chunling Lu et al.

items and a 12-month recall period. The means were significantly different at the 0.99 confidence level, with a *t*-value of 3.43.

In the Bulgaria Living Standards Measurement Study 2001, with a sample size of 2633, the average yearly out-of-pocket health spending generated from the health module with a 1-month recall and five items was 505 leva (95% CI: 440–569). This figure was significantly higher than the 138 leva (95% CI: 128–147) found in the consumption module with a 12-month recall and seven items. We suspect that when questions about health expenditure are fielded in a health module where a respondent has been primed to think about recent health experiences, the estimate may be higher than that resulting from a health-care consumption module. However, we cannot examine this effect directly with the information available.

Effect of survey instrument design

Fig. 3 (available at: http://www.who.int/bulletin/volumes/87/03/08-054379/en/index.html) illustrates the ratio of the percentage of households experiencing catastrophic health spending derived from the single-item measure to that derived from the eight-item measure in the World Health Survey.

The ratio ranges from 0.166 (95% CI: 0.162–0.169) in Slovakia to about 1.965 95% CI: 1.955–1.985) in Uruguay. The observed variation in the ratio suggests that the methods used to collect health expenditure information can significantly confound analyses of the determinants of catastrophic spending and their variations over time. Since reducing catastrophic spending is an important policy objective in several countries, the sensitivity of catastrophic spending to the way information on health expenditures is elicited raises doubts about our capacity to measure the level or trend of these payments.

Discussion

Health expenditure estimates in the same year generated by different surveys can vary greatly.[33] In addition, this paper demonstrates that estimates of household spending on health care are sensitive to the survey instrument design. Usually, a shorter recall period and a longer questionnaire appear to lead to a higher mean estimate of

Table 5. **Recall period and number of questions, Nepal Living Standards Measurement Study, 1997**

	Recall period	
	Past 30 days	**12 months**
Health module (individual)		
Chronic diseases		1
Other illnesses and injury	1	
Consumption module (household)		
Traditional medicines and health services	1	1
Modern medicine and health services (physician fees, hospital charges, and other charges)	1	1

health spending. However, when these survey effects are combined, it is hard to predict which factor – recall period or number of questions – will have the greater effect.

Even the same instrument can generate different response patterns in different populations – a phenomenon known as "differential item functioning".[34,35] The wide variability between countries in the ratio of estimated average spending derived from the single-item question to that derived from the eight-item question in the World Health Survey data indicates that differential item functioning is an important concern. The phenomenon presents major challenges for improving our knowledge of levels and trends in private spending and catastrophic or impoverishing health spending.

The effects of survey design on estimates of spending pose technical challenges for policy discussions about the right mix of private versus public expenditure, as well as for the evaluation of health system performance in developing countries. How can better instruments and estimates of private spending and catastrophic or impoverishing health payments be developed? Progress is needed in three areas. First, new instruments that are less sensitive to the local cultural context and survey design are needed. We recommend that alternative methods be tested in settings where a reasonable approximation of the gold standard measurement of health expenditure is available. This would require selection of validation sites that would enable "true" expenditure information to be obtained, and development and implementation of validation methods to test what kind of survey design can generate estimates closest to the "true" expenditure. In

lower income countries, however, creating a validation environment where "true" expenditure is known may only be possible by identifying all health-care providers, including pharmacies, and recording transactions. Considerable effort and innovation will be needed to create effective validation environments where new instruments can be developed, tested and modified.

Second, any effective new instruments for collecting information on out-of-pocket spending would need to be broadly adopted. This would require substantial efforts to convene stakeholder institutions interested in comparable information on expenditures, such as national statistical offices, The World Bank, WHO and many bilateral donors. Entities that have effectively fostered interest in comparable national health accounts in high-income countries, such as the Organisation for Economic Co-operation and Development, will be critical, as will WHO, to reaching a consensus on dissemination of any new standardized instruments.

Third, non-survey methods may also be helpful in tracking national out-of-pocket spending and other private spending on health (e.g. by nongovernment organizations or private enterprise). These methods may be based on the measurement of a proxy for private health spending (e.g. drug sales, provider surveys that capture both charges and use, tax returns for health providers, human resource data and average salaries, etc.). These methods may be useful additions to surveys but will need to be supplemented by household survey data to track catastrophic or impoverishing health payments.

Comparisons between catastrophic or impoverishing health payments across countries or in the same country

over time must be interpreted with caution, given the extensive evidence of variability in instruments and in differential item functioning for a single instrument. Xu et al. report that 77.2% of the variance in catastrophic health payments could be explained by the fraction of total health expenditure due to out-of-pocket spending, the poverty index and health service utilization.[3] Part of the rest of the variation could be due to measurement error. Further work on strengthening the basis for tracking catastrophic and impoverishing health payments is urgently needed. Meanwhile, the present lack of robust measurement methods should not be an excuse for not addressing the problem of high out-of-pocket spending and families facing financial catastrophe as a result of purchasing health care. ■

Résumé

Limitations des méthodes de mesure des débours directs et des dépenses de santé catastrophiques des ménages

Objectif Etudier l'effet du type d'enquête, et en particulier du nombre de postes de dépense examinés et de la période de rappel, sur les estimations des débours directs et des dépenses catastrophiques en faveur de la santé des ménages.

Méthodes Nous avons utilisé les résultats de deux enquêtes - l'Enquête sur la santé dans le monde et la Living Standards Measurement Study - ayant interrogé les mêmes personnes à propos de leurs dépenses de santé, mais de manières différentes. Les données provenant de l'Enquête sur la santé dans le monde ont servi à comparer les estimations des débours directs annuels moyens des ménages pour la santé, obtenues par une méthode de mesure ne considérant qu'un poste de dépense et par une méthode prenant en compte huit postes. Cette comparaison a été effectuée en déterminant le rapport de la moyenne obtenue par la première méthode de mesure à celle fournie par la seconde méthode. Nous avons également comparé des estimations des dépenses catastrophiques établies à partir de ces deux méthodes de mesure. Nous avons utilisé des données de la Living Standards Measurement Study pour trois pays (Bulgarie, Jamaïque et Népal), correspondant à différentes périodes de rappel et à un nombre variable de catégories de dépenses, et relevées dans le cadre de divers modules d'enquête, pour comparer les estimations des débours directs annuels moyens obtenues par plusieurs méthodes.

Résultats Dans la plupart des pays, un niveau plus faible de désagrégation (c'est-à-dire la prise en compte d'un nombre moindre de postes de dépense) a conduit à une estimation plus basse des dépenses de santé moyennes et une période de rappel plus brève a abouti à une estimation plus élevée. Cependant, lorsque les effets de la désagrégation et de la période de rappel se combinent, il est difficile de prédire quel facteur s'exerce le plus fortement.

Conclusion L'ampleur des débours directs et des dépenses catastrophiques pour la santé des ménages est influencée par le choix de la période de rappel et du nombre de postes de dépense considérés. Il est donc essentiel de définir une méthode pour générer des données valides, fiables et comparables sur les dépenses de santé des privées.

Resumen

Limitaciones de los métodos de medición del gasto sanitario privado, directo y catastrófico

Objetivo Investigar el efecto del diseño de las encuestas, específicamente del número de parámetros y del periodo de rememoración, en las estimaciones del gasto directo (de bolsillo) y los gastos catastróficos de los hogares en salud.

Métodos Utilizamos los resultados de dos encuestas -la Encuesta Mundial de Salud y el Estudio de medición de los niveles de vida- en las que se formulaban preguntas planteadas de distinta manera a las mismas personas sobre los gastos en salud. Los datos de la Encuesta Mundial de Salud se usaron para comparar las estimaciones del gasto directo anual medio en salud obtenidas a partir de un solo parámetro y a partir de ocho parámetros. El resultado se expresó como la razón entre la media obtenida con un solo parámetro y la obtenida midiendo los ocho parámetros. Se compararon asimismo las estimaciones de los gastos catastróficos que arrojaron esos dos métodos de medición. Los datos del Estudio de medición de los niveles de vida de tres países (Bulgaria, Jamaica y Nepal) con distintos periodos de rememoración y distinto número de parámetros en diferentes módulos se usaron para comparar las estimaciones del gasto directo anual medio obtenidas con diversos métodos.

Resultados En la mayoría de los países, un menor grado de desglose (esto es, menos parámetros de valoración) implicaba una estimación menor del gasto sanitario medio, y un periodo de rememoración más corto arrojaba una estimación mayor. No obstante, si se combinan los efectos de la agregación y del periodo de rememoración, resulta difícil predecir cuál de los dos influye más.

Conclusión La magnitud de ambos, los gastos directos y los gastos catastróficos en salud, depende del periodo de rememoración elegido y del número de parámetros empleado. Así pues, es fundamental establecer un método que genere información válida, fiable y comparable sobre el gasto sanitario privado.

ملخص

قصور طرق قياس المصروفات الشخصية والكارثية على الصحة في القطاع الخاص

الهدف: دراسة أثر تصميم المسح الذي يستهدف تقدير ما تدفعه الأسر من مصاريف شخصية (من الجيوب) ومصاريف كارثية على الصحة، ولاسيّما من حيث عدد بنود المسح وزمن التذكّر.

الطريقة: استخدم الباحثون النتائج التي قدمها مسحان هما المسح العالمي للصحة ودراسة قياس مستوى المعيشة، وقد تضمّنت الأسئلة في المسحين اللذين كان لهما نفس الجمهور المستجيب المصروفات الصحية بطرق مختلفة. واستخدم الباحثون المعطيات التي نجمت عن المسح العالمي للصحة لمقارنة التقديرات الخاصة بالمصروفات السنوية الشخصية (من الجيب الشخصي) على الرعاية الصحية، وهذه المعطيات مستمدّة من قياس بند واحد ومن قياس لثمانية بنود. وأجري الحساب للنسبة الوسطية التي حصلنا عليها بقياس بند واحد إلى النسبة الوسطية التي حصلنا عليها بقياس ثمانية بنود. تم مقارنة تقديرات النفقات الباهظة المرتبطة بالقياسين كما استخدموا المعطيات المستمدّة من دراسة قياس مستوى المعيشة التي شملت ثلاثة

بلدان هي بلغاريا وجامايكا ونيبال مع فترات مختلفة للتذكر وأعداد متفاوتة للبنود وردت في نماذج أو كتيبات مختلفة لمقارنة تقديرات المتوسطّ السنوي للمصروفات من الجيب الشخصي وباستخدام طرق مختلفة.

الموجودات: في معظم البلدان أدى المستوى المنخفض من التصنيف (العدد القليل من البنود) إلى تقديرات منخفضة للمصروفات الوسطية على الصحة، كما أدى الزمن الأقصر للتذكّر إلى تقديرات أكبر. ومع ذلك فإنه عند جمعِ تأثيرات التجميع مع زمن التذكّر معاً يصبح من الصعب التوقُّع أي من هذين العاملين هو الأكثر تأثيراً.

الاستنتاج: يتأثر مدى كل من المصروفات الشخصية والكارثية باختيار زمن التذكّر وبعدد البنود ومن هنا كان وضع الطريقة التي تسمح بالحصول على معلومات صحيحة وموثوقة وممكن مقارنتها بغيرها أمراً هاماً في تقديرات المصارف في القطاع الخاص على الصحة.

References

1. *The world health report 2006: working together for health*. Geneva: World Health Organization; 2006.
2. *The world health report 2000: health systems: improving performance*. Geneva: World Health Organization; 2000.
3. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJL. Household catastrophic health expenditure: a multi-country analysis. *Lancet* 2003;362:111-7. PMID:12867110 doi:10.1016/S0140-6736(03)13861-5
4. Xu K, Klavus J, Kawabata K, Evans DB, Hanvoravongchai P, Ortiz de Iturbide JP, et al. Household health system contributions and capacity to pay: definitional, empirical and technical challenges. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization; 2003.
5. Wagstaff A, van Doorslaer E. Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993–1998. *Health Econ* 2003;12:921-34. PMID:14601155 doi:10.1002/hec.776
6. Xu K, Klavus J, Aguilar-Rivera AM, Carrin G, Zeramdini R, Murray CJL. Summary measures of the distribution of household financial contributions to health. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization; 2003.
7. van Doorslaer E, O'Donnell O, Rannan-Eliya RP, Somanathan A, Adhikari SR, Garg CC, et al. Effect of payments for health care on poverty estimates in 11 countries in Asia: an analysis of household survey data. *Lancet* 2006;368:1357-64. PMID:17046468 doi:10.1016/S0140-6736(06)69560-3
8. Su TT, Kouyaté B, Flessa S. Catastrophic household expenditure for health care in a low-income society: a study from Nouna District, Burkina Faso. *Bull World Health Organ* 2006;84:21-7. PMID:16501711 doi:10.2471/BLT.05.023739
9. Liu Y, Rao K, Hsiao WC. Medical expenditure and rural impoverishment in China. *J Health Popul Nutr* 2003;21:216-22. PMID:14717567
10. Kawabata K, Xu K, Carrin G. Preventing impoverishment through protection against catastrophic health expenditure. *Bull World Health Organ* 2002; 80:612. PMID:12219150
11. Article 90 in Law of the fourth economic, social and cultural development plan of the Islamic Republic of Iran, 2005–2009. Tehran: Management and Planning Organization; 2004.
12. Pannarunothai S, Patmasiriwat D, Srithamrongsawat S. Universal health coverage in Thailand: ideas for reform and policy struggling. *Health Policy* 2004;68:17-30. PMID:15033549 doi:10.1016/S0168-8510(03)00024-1
13. Frenk J, Gonzalez-Pier E, Gomez-Dantes O, Lezana MA, Knaul FM. Comprehensive reform to improve health system performance in Mexico. *Lancet* 2006;368:1524-34. PMID:17071286 doi:10.1016/S0140-6736(06)69564-0

14. Gonzalez-Pier E, Gutierrez-Delgado C, Stevens G, Barraza-Llorens M, Porras-Condey R, Carvalho N, et al. Priority setting for health interventions in Mexico's system of social protection in health. *Lancet* 2006;368:1608-18. PMID:17084761 doi:10.1016/S0140-6736(06)69567-6
15. Lozano R, Soliz P, Gakidou E, Abbott-Klafter J, Feehan DM, Vidal C, et al. Benchmarking of performance of Mexican states with effective coverage. *Lancet* 2006;368:1729-41. PMID:17098091 doi:10.1016/S0140-6736(06)69566-4
16. Knaul FM, Arreola-Ornela H, Mendez-Carniado O, Bryson-Cahn C, Barofsky J, Maguire R, et al. Evidence is good for your health system: policy reform to remedy catastrophic and impoverishing health spending in Mexico. *Lancet* 2006;368:1828-41. PMID:17113432 doi:10.1016/S0140-6736(06)69565-2
17. Gakidou E, Lozano R, Gonzalez-Pier E, Abbott-Klafter J, Barofsky JT, Bryson-Cahn C, et al. Assessing the effect of the 2001-06 Mexican health reform: an interim report card. *Lancet* 2006;368:1920-35. PMID:17126725 doi:10.1016/S0140-6736(06)69568-8
18. Sepulveda J, Bustreo F, Tapia R, Rivera J, Lozano R, Olaiz G, et al. Improvement of child survival in Mexico: the diagonal approach. *Lancet* 2006; 368:2017-27. PMID:17141709 doi:10.1016/S0140-6736(06)69569-X
19. Organisation for Economic Co-operation and Development. *A system of health accounts*. Paris: OECD; 2000.
20. *Guide to producing national health accounts: with special applications for low-income and middle-income countries*. Geneva: World Health Organization; 2003.
21. Winter J. Response bias in survey-based measures of household consumption. *Econ Bull* 2004;3:1-12.
22. Neter J, Waksberg J. A study of response errors in expenditures data from household interviews. *J Am Stat Assoc* 1964;59:18-55. doi:10.2307/2282857
23. Neter J. Measurement errors in reports of consumer expenditures. *J Mark Res* 1970;7:11-25. doi:10.2307/3149502
24. Eisenhower D, Mathiowetz NA, Morganstein D. Recall error: sources and bias reduction techniques. In: Biemer P, Sudman S, Groves RM, eds. *Measurement error in surveys*. New York: Wiley and Sons; 1991.
25. Beckett M, DaVanzo J, Sastry N, Panis C, Peterson C. The quality of retrospective data: an examination of long-term recall in a developing country. *J Hum Resour* 2001;36:593-625. doi:10.2307/3069631
26. Battistin E, Miniaci R, Weber G. What do we learn from recall consumption data? *J Hum Resour* 2003;38:354-85. doi:10.2307/1558748
27. Browning M, Crossley TF, Weber G. Asking consumption questions in general purpose surveys. *Econ J* 2003;113:F540-67. doi:10.1046/j.0013-0133.2003.00168.x

28. Deaton A, Grosh M. Consumption. In: Grosh M, Glewwe P, eds. *Designing household survey questionnaires for developing countries: lessons from 15 years of the Living Standards Measurement Study*. Washington, DC: The World Bank; 2000.

29. World Health Survey. Geneva: World Health Organization. Available from: http://www.who.int/healthinfo/survey/en/index.html [accessed on 16 January 2007].

30. Living Standards Measurement Study. Washington, DC: The World Bank. Available from: http://www.worldbank.org/lsms/ [accessed on 16 January 2007].

31. Lanjouw JO, Lanjouw P. How to compare apples and oranges: poverty measurement based on different definitions of consumption. *Rev Income Wealth* 2001;47:25-42. doi:10.1111/1475-4991.00002

32. Pradhan M. *Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia* [working paper]. Amsterdam: Free University; 2001.

33. Xu K, Ravndal F, Evans D, Carrin G. *Assessing the reliability of household expenditure data: results of the World Health Survey*. Geneva: World Health Organization; 2007 (Discussion Paper 5).

34. Holland PW, Wainer H. *Differential item functioning*. Hillsdale, NJ: Erlbaum; 1993.

35. King G, Murray CJL, Salomon JA, Tandon A. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am Polit Sci Rev* 2004;98:191-207.

Table 2. **Description of data from the World Health Survey**

Country code	Country	No. of households surveyed	Items responded to (%)	No. of observations used in Fig. 1, Fig. 2 and Fig. 3
ARE	United Arab Emirates	1 183	99.75	1 182
BFA	Burkina Faso	4 945	97.51	4 942
BGD	Bangladesh	5 942	93.44	5 936
BIH	Bosnia and Herzegovina	1 031	99.71	1 031
CHN	China	3 994	99.97	3 993
CIV	Côte d'Ivoire	3 245	97.94	3 224
COG	Congo	3 070	81.11	3 021
COM	Comoros	1 835	95.80	1 835
CZE	Czech Republic	949	98.52	949
DOM	Dominican Republic	5 027	90.19	5 017
ECU	Ecuador	5 645	81.98	4 717
EST	Estonia	1 021	99.12	1 018
ETH	Ethiopia	5 090	97.01	5 087
GEO	Georgia	2 947	93.38	2 756
GHA	Ghana	4 159	94.54	5 151
HRV	Croatia	993	99.70	993
IND	India	10 683	93.47	10 578
KEN	Kenya	4 639	95.19	4 621
LAO	Lao People's Democratic Republic	4 989	98.00	4 974
LKA	Sri Lanka	6 805	98.93	6 798
LVA	Latvia	929	92.14	929
MAR	Morocco	5 000	100.00	5 000
MEX	Mexico	38 746	100.00	38 746
MLI	Mali	5 195	82.21	5 179
MRT	Mauritania	3 844	98.31	3 830
MWI	Malawi	5 545	95.58	5 535
MYS	Malaysia	6 145	98.29	6 133
NAM	Namibia	4 377	97.05	4 359
NPL	Nepal	8 822	98.48	8 801
PAK	Pakistan	6 502	98.11	6 495
PHL	Philippines	10 083	99.95	9 634
RUS	Russian Federation	4 426	99.89	4 411
SEN	Senegal	3 458	93.09	3 438
SVK	Slovakia	2 532	99.37	1 821
SWZ	Swaziland	3 070	100.00	2 904
TCD	Chad	4 866	95.60	4 858
TUN	Tunisia	5 199	97.42	5 181
UKR	Ukraine	2 811	99.72	2 803
URY	Uruguay	2 986	99.73	5 972
VNM	Viet Nam	4 174	83.66	4 173
ZAF	South Africa	2 601	89.35	2 596
ZMB	Zambia	4 165	91.50	4 161
ZWE	Zimbabwe	4 264	95.50	4 237

Fig. 1. **Annual out-of-pocket health spending: ratio of average derived from single-item measure to average derived from eight-item measure, World Health Survey, 2003**
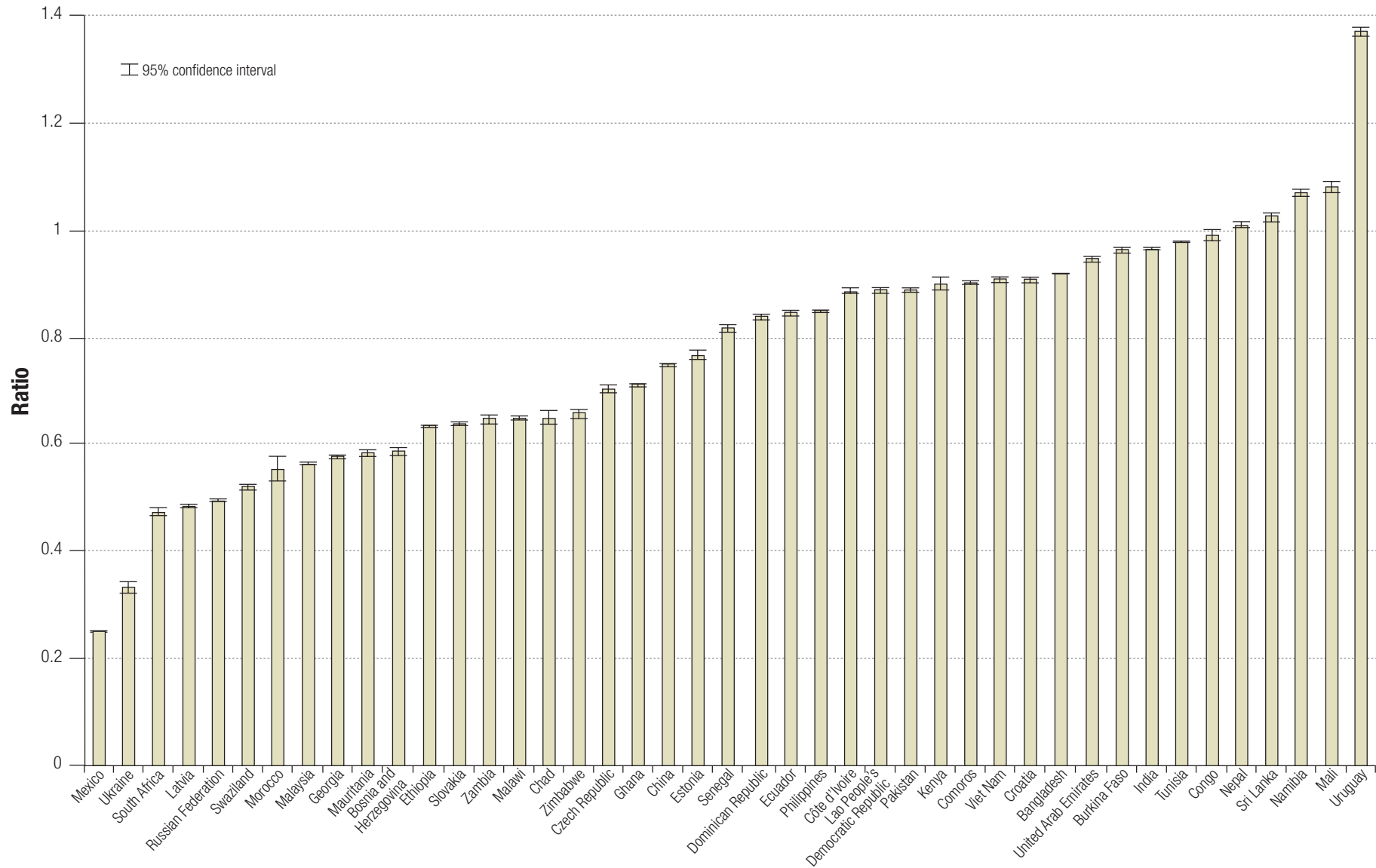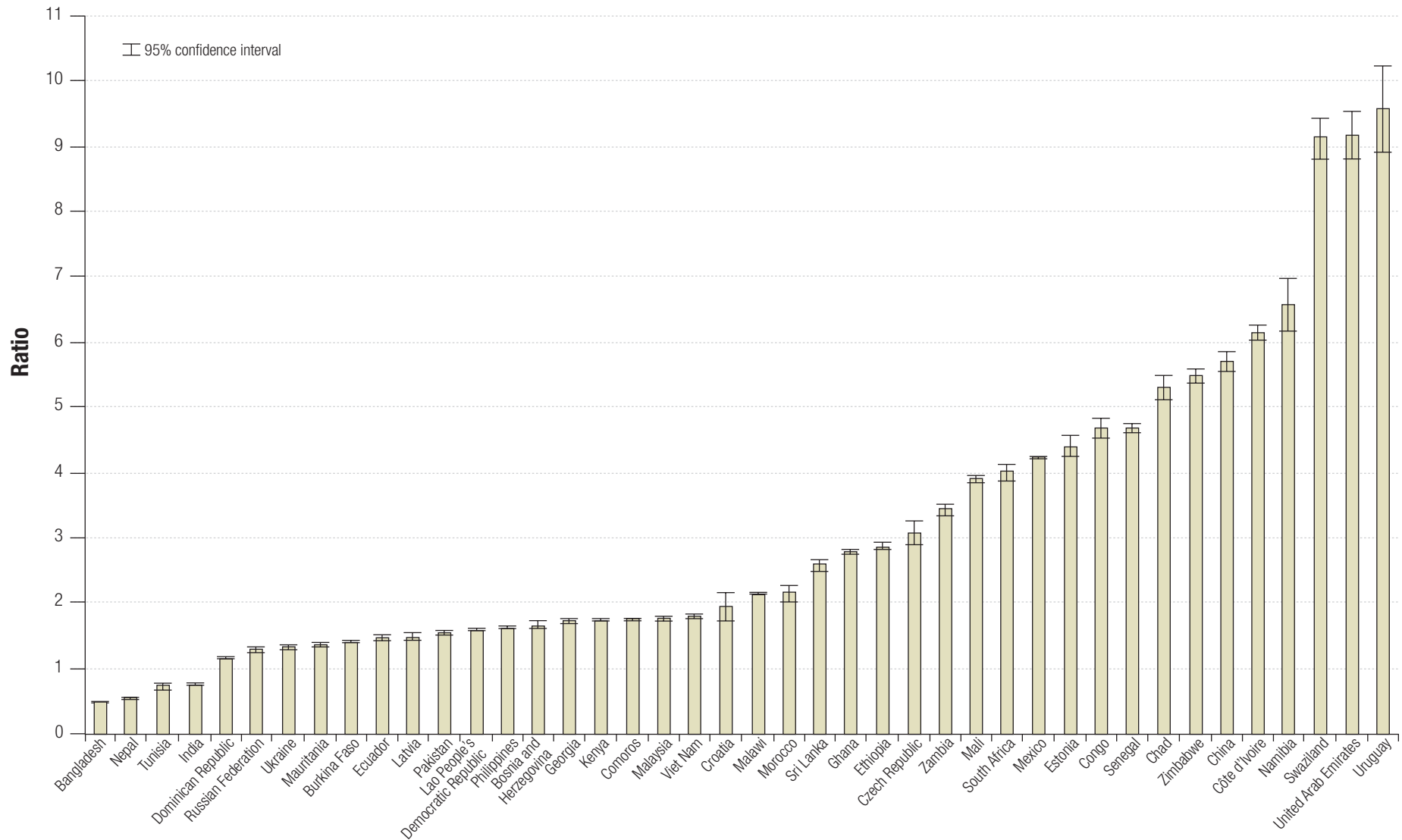
Research
Chunling Lu et al.
Measuring private health expenditures

Fig. 2. **Annual out-of-pocket health spending on hospitalization: ratio of average derived from single-item measure to average derived from eight-item measure, World Health Survey, 2003[a]**

[a] Slovakia is not included in the figure because its ratio and confidence intervals are high (16.43+/-13.37) so excluded to show more clearly the differentials for other countries.

C

Fig. 3. **Catastrophic health spending: ratio of average derived from single-item measure to average derived from eight-item measure, World Health Survey, 2003**