Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine
Health Policy and Planning 2012;**27**:76–83
doi:10.1093/heapol/czr004

How to do (or not to do)...Assessing the impact of a policy change with routine longitudinal data

Mylene Lagarde

Health Economics and Financing Programme, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. Tel: +44 20 7927 2090. E-mail: Mylene.Lagarde@lshtm.ac.uk

Accepted 15 October 2010

A lack of good quality evidence on the effect of alternative social policies in low- and middle-income countries has been recently underlined and the value of randomized trials increasingly advocated. However, it is also acknowledged that randomization is not always feasible or politically acceptable. Analyses using longitudinal data series before and after an intervention can also deliver robust results and such data are often reasonably easy to access.

Using the example of evaluating the impact on utilization of a change in health financing policy, this article explains how studies in the literature have often failed to address the possible biases that can arise in a simple analysis of routine longitudinal data. It then describes two possible statistical approaches to estimate impact in a more reliable manner and illustrates in detail the more simple method. Advantages and limitations of this quasi-experimental approach to evaluating the impact of health policies are discussed.

Keywords Evaluation, time-series, health policy change, user fees

Introduction

Recent articles regarding the evaluation of social policies in low- and middle-income countries (Duflo *et al*. 2004; Savedoff *et al*. 2006) have emphasized the weakness of evidence concerning their effectiveness. Unlike testing the efficacy of a drug, investigating the success of a particular policy is more complex, due to the variety of possible causes of any observed trend. Randomized experiments are the gold standard by which effectiveness is measured in clinical disciplines, but they can be logistically difficult to implement when it comes to social sectors (Ranson *et al*. 2006). Economic obstacles (impact evaluations are costly and labour-intensive) or political constraints (it is not always possible to give services to some communities and not to others) can also prevent the use of randomized controlled trials. Finally, these experiments may be perceived as overly burdensome and time-consuming while changes in policies are sometimes driven by shorter timeframes or local political agendas.

There are alternative, less demanding study designs that can be used in an attempt to improve the quality of information for decision-makers. In its standard inclusion criteria (Cochrane Effective Practice and Organisation of Care Review Group 2002), the Effective Practice and Organisation of Care Group (EPOC) of the Cochrane Collaboration recognizes three different types of study design. In addition to randomized experiments, two types of quasi-experimental designs (Cook and Campbell 1979) are considered for inclusion: controlled before and after studies and interrupted time-series (ITS) studies (Grimshaw *et al*. 2003).

ITS studies use routine data collected at equally spaced intervals of time before and after an intervention, and do not necessarily require a control site, which makes it a much more practical option in many cases: 'Interrupted time-series studies can provide a robust method of measuring the effect of an intervention when randomization or identification of a control group are impractical' (Grimshaw *et al*. 2003). This approach has been used to assess the consequences of a variety of policy issues in various fields, such as environmental policies (Box and Tiao 1975), financial economics (Ho and Wan 2002) and in some cases, health policies (van Driel *et al*. 2008; Chan *et al*. 2009; Zhang *et al*. 2009).

Nevertheless, this study design is not yet widely known in the health policy community and appears to be particularly underused in the monitoring of health policy changes in low- and middle-income countries despite some appealing features. When interventions have been implemented without the opportunity to conduct a baseline survey before the intervention, administrative routine data are still often available, even for several years before the intervention. This approach allows researchers to evaluate the impact of interventions that have been implemented at a precise point in time, and were expected to have effects on outcomes captured by the health information system (for instance, number of vaccinations, number of outpatient/inpatient visits, number of cases of fever, number of deliveries, etc.). Such interventions typically include human resource interventions (salary increase), management reforms (contracting in health service provision, decentralization, etc.) or changes to health financing.

The objective of this article is to encourage the evaluation of health policy changes through adequate analysis when only retrospective longitudinal data obtained before and after an intervention are available. First, the article presents the potential biases that may arise if longitudinal data are analysed too simplistically. Secondly, it describes what statistical analysis can be performed relatively easily by the analyst. This article takes a hypothetical example of changes in health financing, but the points made in the paper can apply to the analysis of other types of health policies.

Issues in the analysis of longitudinal data

To date, despite a significant body of literature using routine longitudinal data to assess the impact of health financing policy change on utilization (Moses *et al*. 1992; Mbugua *et al*. 1995; Wilkinson *et al*. 2001; Ridde 2003; Burnham *et al*. 2004; Nabyonga *et al*. 2005), the evidence arising from these analyses is prone to bias, in particular due to the nature of the statistical analyses performed (Lagarde and Palmer 2008). In particular, very few studies (Collins *et al*. 1996; Wilkinson *et al*. 2001) have attempted to account for the specific properties of time-series data and the risks arising from these. Many studies using longitudinal data to assess the impact of user fee policy change (Moses *et al*. 1992; Kipp *et al*. 2001; Meuwissen 2002; Ridde 2003; Akashi *et al*. 2004; Burnham *et al*. 2004; Nabyonga *et al*. 2005) have limited their analysis to the comparison of average utilization data before and after the change of policy. Box and Tiao (1975) have emphasized the irrelevance of procedures such as evaluation of Student's t to compare means, as the particular characteristics of time-series—non-stationarity, seasonality and auto-correlation (see Table 1)—may lead to biased results.

Non-stationarity relates to the series exhibiting one or more secular trends, implying that the mean and variance of the data series can change over time for reasons other than the effect of the policy change. The analysis of time-series often starts with de-trending the series, or transforming the data to leave a stationary data set for analysis. Alternatively, one can introduce a variable that captures the structural trend. Seasonality refers to periodic fluctuations. For example, in a malarial area, curative consultations are likely to peak during the malaria transmission season then decline afterwards. Finally, autocorrelation of a time-series describes the correlation between values of the series at different points in time. The presence of autocorrelation can invalidate the use of simple statistical tools such as ordinary least squares (OLS) regressions. Indeed, the presence of autocorrelation violates the OLS assumption that the error terms are uncorrelated. While this does not bias the estimates of the regression coefficients, the standard errors tend to be underestimated, possibly leading to biased conclusion about the statistical significance of some coefficients.

Some simple graphic illustrations provide an insight into some of the risks of bias that can arise from using longitudinal data too simply. Figure 1, created from hypothetical data, illustrates four potential mis-readings of results that can arise from the simple comparison of means of longitudinal data

Table 1 Summary of main issues in the use of longitudinal data

Potential source of bias	Definition	Resulting problem
Non-stationarity	There is a 'natural' trend in the data (e.g. upward sloping), independent from other events.	The mean value (and variance) of the outcome of interest naturally changes over time (e.g. natural increase), independently from any event such as the studied intervention.
Auto-correlation	Values at one point in time are statistically correlated with past values of the data. Typically, adjacent data points in time are more likely to be close to each other than points that are further from each other (this is called first-order correlation—a point at a date t is correlated with the point at date t-1). Another form of correlation might be linked to seasonality: the level of an outcome over a particular season might be correlated to the level over the same season the year before.	One of the assumptions in OLS regressions is that the error terms associated with each observation are uncorrelated. If this assumption is violated, estimates of the standard errors of the coefficients in the regression will be biased. This can lead to incorrect results on the statistical significance of coefficients.
Seasonality	There are some regular (expected) changes in the outcome due to seasonal effects (e.g. increase in utilization due to the seasonality of some diseases).	The level of the outcome varies at certain times (seasons) independently of other factors; not controlling for these changes in levels can blur the real effect of an intervention.

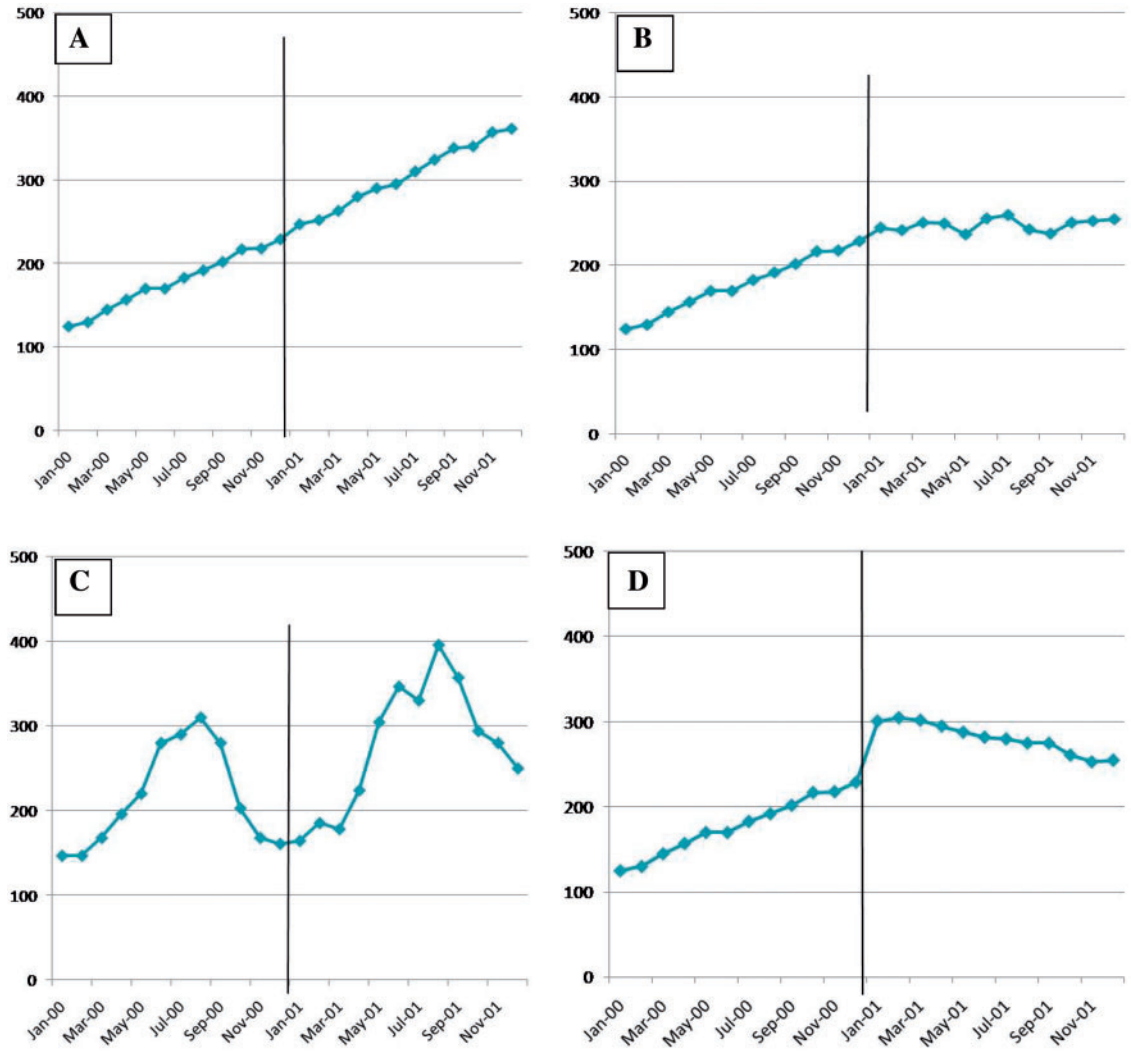

Figure 1 Potential mis-readings of results arising from simple comparison before and after of means of longitudinal data

before and after an intervention. In each of the four cases, the average computed after the hypothetical intervention (represented by the vertical line) is greater than the average computed before. Some would conclude that the intervention was successful in increasing the outcome of interest. However, in each of these four examples, that would be a wrong conclusion to draw from the data.

In the first case (A), there is clearly an upward trend in the data series, which started before the intervention. This is a situation where the time-series is non-stationary: the upward trend automatically changes the mean over the series over time. In such a situation, the upward trend by itself leads to a higher mean after the intervention compared with before, regardless of the effects of the intervention. In the second case (B), the upward trend that had started before seems to have stopped after, to be replaced by a flatter trend (the series is non-stationary before the intervention, but stationary after). However, given the higher overall level of utilization, a comparison of the two mean outcomes, before and after, would erroneously show that the situation had improved due to the policy intervention. The third example (C) provides an example of how seasonal effects (i.e. seasonality in the data) can blur the interpretation of average outcomes. Although the average after would be greater than the average before, this could mainly be due to a stronger seasonal effect in the post-intervention period. Finally, the last example (D) shows a situation where there would be a steep increase in the outcome right after the policy change, but the trend in outcome following the intervention is downward sloping, suggesting some concerns about the sustainability of the effect. This again, however, would give an overall increase before and after the policy change, although the intervention may just have created a change of direction in the trends of the data.

These simple examples also illustrate the fact that comparing mean outcomes before and after is very limiting given the wealth of information contained in a data series such as these. The longer the series, the more precise information one can obtain on the trend of the outcome before and after the policy change, the possible presence of seasonal effects and whether or not the policy has a lasting effect on the outcome.

The rest of this article describes two methods that can take into account the problems presented in Figure 1, and more successfully detect the effect of an intervention.

Analysing the effect of a policy change with longitudinal data

As explained above, ITS studies refer to a general study design that relies on the collection of longitudinal data before and after an intervention occurred. Having showed how a simple analysis of the data could be misleading, I now describe two approaches to analysing correctly such longitudinal data, and provide a detailed description of the more simple approach.

Using ARIMA modelling

McDowall *et al*. (1980) suggest an approach to analysing longitudinal data using Auto-Regressive Integrated Moving Average (ARIMA) modelling based on the Box–Jenkins methodology (Box and Jenkins 1970). This approach however has several difficulties and limitations in the context of health systems in low- and middle-income countries.

An ARIMA model is used to account for all specificities of time-series data by capturing any underlying systematic time-series patterns in the data—non-stationarity or seasonality—and accounting for auto-correlation of the series. This was a method developed in econometrics for forecasting the future behaviour of an outcome of interest, based on its past behaviour (Hamilton 1994).

The first step of the method is to specify the best and most parsimonious ARIMA model fitting the outcome series y_t *before the intervention*. Once that ARIMA process is identified, it is used to model the *entire dataset (pre- and post-intervention)* augmented by an intervention component. That intervention component can take several forms according to the assumption(s) researchers want to test in the modelling of the intervention effect. Examples of intervention shapes include an abrupt and permanent change modelled with a step function, or an abrupt and temporary change modelled by a pulse function. Other examples can be found in the relevant literature (McDowall *et al*. 1980).

The main difficulty of this approach to analysing ITS data resides in the command of very sophisticated statistical skills, in particular those needed to identify the best-fitting ARIMA model. For an adequate introduction to the analysis of time-series, see relevant textbooks such as Brockwell and Davis (2002) or Hamilton (1994). In addition, ARIMA models are usually considered robust for a long time-series, featuring at least 100 points. Unfortunately, in low- and middle-income countries, the health information systems rarely provide information of this length and consistency. Furthermore, based on econometric techniques that have been used more in a predictive than an explanatory approach, this approach is less flexible and not necessarily fit for purpose to evaluate the effect of a policy. Finally, this technique is likely less known by public health evaluators with other than economist backgrounds.

Using a segmented linear regression

The alternative method of analysis uses more simple linear regression techniques. This approach, sometimes called segmented regression analysis (Ramsay *et al*. 2001; Wagner *et al*. 2002), controls for secular trends and can also adjust for potential serial correlation of the data.

The specification of the linear regression to be analysed is:

$$Y_t = \beta_0 + \beta_1^* \text{time} + \beta_2^* \text{intervention} + \beta_3^* \text{postslope} + \varepsilon_t \quad (1)$$

Where Y_t is the outcome variable at time t; time is a continuous variable indicating time from the start of the study up to the end of the period of observation; intervention is coded 0 for pre-intervention time points and 1 for post-intervention time points and postslope is coded 0 up to the last point before the intervention phase and coded sequentially from 1 thereafter (see Table 2).

In this model, β_0 captures the baseline level of the outcome at time 0 (beginning of the period); β_1 estimates the structural trend or growth rate in utilization, independently from the intervention; β_2 estimates the immediate impact of the intervention or the change in level in the outcome of interest after the intervention; and β_3 reflects the change in trend, or growth rate in outcome, after the intervention.

An alternative coding that highlights the trends before and after the intervention is:

$$Y_t = \gamma_0 + \gamma_1^* \text{preslope} + \gamma_2^* \text{intervention} + \gamma_3^* \text{postslope} + \varepsilon_t \quad (2)$$

where preslope is a continuous variable indicating time from the start of the study up to the beginning of the intervention (see Table 2). In this model, γ_1 estimates the structural trend in the outcome before the intervention, while γ_3 captures the trend after the intervention. The other two coefficients γ_0 and γ_2 remain equivalent to β_0 and β_2. To measure the *change* in trend, equivalent to coefficient β_3 in equation (1), the analyst can subtract $\gamma_3 - \gamma_1$.

The coding of equation (1) is preferred as it allows the analyst to directly test the secular trend in the data, and the change in trend caused by the intervention.

Furthermore, it is important to control for auto-correlation in the data series. Two approaches exist to do so. First, the standard econometric approach is to perform a Durbin–Watson test to test the presence of first-order auto-correlation (a value around 2 indicates no sign of auto-correlation). If auto-correlation is detected, a generalized least squares estimator, such as the Prais–Winsten method (Judge *et al*. 1985), should be used to estimate the regression—this is easily done in STATA by using the command 'prais' instead of 'reg' to perform the regression. Alternatively, in the multilevel regression context familiar to epidemiologists, one would use random intercepts and random coefficients, and test a first-order autoregressive (or some other available) level one covariance structure (Singer and Willett 2002).

Data layout for a segmented linear regression

To illustrate the analysis, data from a health district in Zambia where fees were removed in April 2006 are used. The outcome of interest (called 'outcome' in Table 2) is the number of monthly outpatient consultations. To perform either one of the two analyses presented in equation (1) and equation (2),

Table 2 Dataset and variables used in the analysis

(Actual) Date	outcome	time	intervention	preslope	postslope
Jan-05	5361	1	0	1	0
Feb-05	4525	2	0	2	0
Mar-05	4620	3	0	3	0
Apr-05	4709	4	0	4	0
May-05	4632	5	0	5	0
Jun-05	4743	6	0	6	0
Jul-05	5118	7	0	7	0
Aug-05	5245	8	0	8	0
Sep-05	4857	9	0	9	0
Oct-05	4075	10	0	10	0
Nov-05	3983	11	0	11	0
Dec-05	3953	12	0	12	0
Jan-06	4602	13	0	13	0
Feb-06	5418	14	0	14	0
Mar-06	6302	15	0	15	0
Apr-06	5750	16	1	15	1
May-06	5773	17	1	15	2
Jun-06	5900	18	1	15	3
Jul-06	6922	19	1	15	4
Aug-06	7064	20	1	15	5
Sep-06	6160	21	1	15	6
Oct-06	7008	22	1	15	7
Nov-06	8202	23	1	15	8
Dec-06	8022	24	1	15	9
Jan-07	6059	25	1	15	10
Feb-07	6189	26	1	15	11
Mar-07	6852	27	1	15	12
Apr-07	5991	28	1	15	13
May-07	6115	29	1	15	14
Jun-07	5807	30	1	15	15
Jul-07	5803	31	1	15	16
Aug-07	6607	32	1	15	17
Sep-07	7094	33	1	15	18
Oct-07	6699	34	1	15	19
Nov-07	6614	35	1	15	20
Dec-07	5772	36	1	15	21

the analyst has to create the necessary independent variables ('time', 'intervention', 'preslope' and/or 'postslope') as presented in Table 2. The analysis can then be performed in STATA or any other statistical package, with the outcome variable as the dependent variable in the regression.

Data analysis and interpretation

Using STATA, two models were estimated based on equation (1) and the dataset presented in Table 2. Model 1 is the simple linear regression not adjusting for first-order auto-correlation (despite a Durbin–Watson statistics of 1.23 suggesting its presence) and Model 2 is the same model using a Prais–Winsten estimator that corrects for data auto-correlation (Table 3).

Table 3 Results of two segmented linear regression models

Independent variables	Coefficient	Standard Error	*P*-value
Model 1 (no correction for auto-correlation)			
Constant β_0	4658.48***	378.22	0.000
Secular trend β_1	18.88	41.60	0.653
Change in level β_2	1603.33***	465.09	0.001
Change in trend β_3	−23.40	48.58	0.633
Model 2 (correcting for first-order auto-correlation)			
Constant β_0	4513.30***	658.35	0.000
Secular trend β_1	73.68	67.29	0.282
Change in level β_2	415.13	631.38	0.516
Change in trend β_3	−50.59	88.94	0.573

Note: ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

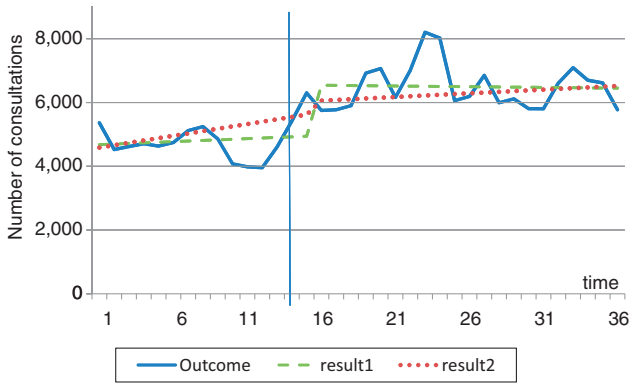

Figure 2 Raw data series of the outcome of interest, and the fitted results; result 1 with no adjustment for first-order auto-correlation, and result 2 correcting for data auto-correlation

Results from Model 1 indicate that at the beginning of the period of observation, there were on average 4658 consultations in the district. There was no significant month-to-month change in the number of consultations, either before or after the intervention (coefficient β_1 and β_3 are not significant). Immediately after the intervention Model 1 suggests that the number of consultations increased suddenly and significantly by about 1603 consultations per month. However, correcting for auto-correlation in the data, Model 2 no longer shows a significant immediate effect of the intervention. These results demonstrate once more the importance of accounting for the properties of time-series in order to interpret the data correctly.

Figure 2 presents the raw data series of the outcome of interest, and the fitted results (dotted lines) obtained with this method. In the graph the outcome of interest and the fitted values of the model have been plotted against the time variable.

The results of the analysis can also be used to calculate the absolute and relative effects of the intervention, for any date after the intervention (e.g. 6, 12 and 18 months). For example, to calculate $\widehat{Y}_{28}$, the expected number of consultations in the district 12 months after the intervention (at month 28), we

simply plug the values of the explanatory variables at month 28 into the estimated regression equation (1):

$$\hat{Y}_{28} = \hat{\beta}_0 + \hat{\beta}_1 \times 28 + \hat{\beta}_2 \times 1 + \hat{\beta}_2 \times 1 + \hat{\beta}_3 \times 13$$

In the absence of a control group, the analyst can also use the results of the regression to fabricate a counterfactual and estimate what would have been the level of the outcome of interest without the intervention and its effects:

$$\hat{Y}_{28}^{NO} = \hat{\beta}_0 + \hat{\beta}_1 \times 28$$

Using the estimated coefficients of Model 2 in Table 3, we find here that 1 year after the intervention there was an average of 6334 consultations per month, while without the intervention the model predicted that health services utilization would have reached an average of 6576 monthly consultations. Therefore, the absolute effect of the intervention was a (non-significant) decrease by about 242 monthly consultations, and the relative effect was a decrease of 3.6% in the number of consultations.

Using the lower and upper bounds of the 95% confidence intervals around the coefficients, it is possible to estimate the 95% confidence interval of the intervention effect.

More refinements could be added to the regression specification. For example, indicator variables could be added to control for particular outliers in the data series (Table 4). Here, the analyst could know that there was a shortage of drugs between October and December 2005 (months 10 to 12) and a disease outbreak in November–December 2006 (this is not the case here). Indicator variables controlling for these two circumstantial events would take the value 1 over the period and 0 otherwise. This allows a refinement of the intervention effects, which can then be recalculated with new values of estimated coefficients of equation (1).

This description provides a basic framework to evaluate the effect of an intervention. This is a simplified approach to the repeated measurement analysis proposed in the more general multilevel framework (Snijders and Bosker 1999; Singer and Willett 2002; Fitzmaurice *et al*. 2004). This framework allows time-series analysis by individual health facilities or areas and can incorporate additional information (typically through facility or area-level covariates), making the assessment of the intervention more robust.

Table 4 Results of a segmented linear regression model with controls for outliers

Independent variables	Coefficient	Standard Error	*P*-value
Constant β_0	4588.91***	281.42	0.000
Secular trend β_1	60.02*	32.39	0.074
Change in level β_2	726.47*	370.22	0.059
Change in trend β_3	−51.37	37.92	0.186
outlier1	−1193.22***	337.31	0.001
outlier2	1788.45***	364.91	0.000

Note: ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

Discussion

Because it only requires relatively accessible data—usually routinely collected by health information systems—segmented regression analysis is a practical approach to assessing the impact of a health policy change in low- and middle-income health systems.

In addition to the results of the regressions, which underline whether or not the intervention was found to be associated with a significant change in the outcome of interest, predicting the outcome levels with and without the policy intervention at regular intervals of time can help the researcher convey the results of the analysis to policy-makers. By presenting relative changes in outcomes between chosen dates after and before the intervention, one can better highlight the issue of the durability of any effect. Indeed, an intervention can produce a steep increase in a particular outcome but then be followed by a gradual downward trend.

Relying on multiple observations over time, before and after an intervention, enables the researcher to control for pre-existing trends and can display learning effects that will alter the impact of a policy change over time. For example, a change in management in health facilities could have immediate quality effects that will only be perceived gradually by the patients who will then change their behaviours (learning effect). In contrast, controlled before and after studies (or even randomized trials) that rely on cross-sectional data can fail to detect the dynamics of the impact of an intervention over time. The sustainability of the impact of a policy change is as important as its initial impact. This would not be possible to assess with simple analyses of longitudinal data or cross-sectional data.

Although this approach requires a good understanding of quantitative tools, it is much less complex than the (ARIMA) time-series approach and it provides more transparent and robust estimates of policy impact than the basic approach often undertaken.

Researchers should be encouraged to use this technique, whilst also keeping in mind the following caveats.

First, it is difficult to infer causality between an observed pattern and the intervention (Mohr 1995; Shadish *et al*. 2002). The analysis enables one to determine whether there was a systematic shift in the target variable time series at and after a given time point, but it does not clearly demonstrate the causal determinants of that shift. Because health policy changes occur in complex health systems, concomitant reforms or events are likely to have a direct or indirect influence on the outcome of interest. For example, a change in health financing such as removing user charges can be accompanied by supply-side reforms affecting indirectly the demand for services (e.g. improvement in management and delivery of health services and increase in human resources). Such policy change can also take place in the middle of broader changes external to the health sector (e.g. economic crisis, currency devaluation) that will have an impact on households' behaviours. To interpret the results and discuss alternative explanations, other (district- or facility-level) explanatory variables can be added to the model. Yet, it is often a challenge to obtain longitudinal data about possibly influential local events (not necessarily limited to the health sector) for interpretation and discussion of alternative

explanations. Another way to overcome the confounding effect of other reforms or events is to conduct the same analysis in a control site. The underlying logic is that trends in intervention areas would have been similar to trends in control areas, had there been no intervention. To provide reliable results, this approach should use closely matched sites, showing structural similarities, as well as dynamic ones. This means that control and intervention sites should not only share similar socio-economic and health system characteristics, but they should display similar utilization patterns before the intervention. Such examples of controls will only rarely be available. Another way to complement such quantitative studies is to use qualitative research to look at the extent to which the intervention of focus was actually implemented (and if not, the problems that were incurred) (Mills *et al*. 2008).

Another possible drawback of the method can stem from the limitations of the data that it relies upon. Firstly, the results of time-series regression studies can vary sharply depending on which sets of time points are used. This issue is particularly sensitive when the sample size is small, which is usually the case in low- and middle-income countries given the scarcity of reliable routine data. The shorter the time-series, the more it will be subject to short-term changes in the target variable and the more likely the analysis is to miss long-term patterns. Secondly, the advantage of relying on easily available data can also become a drawback. Routine data may not always be very complete: health facilities may have an incentive to misreport if their funding is dependent on their activity, and standards of data can change over time. In particular, register keeping and reporting might typically change together with a policy change or an intervention, and such changes in 'instrumentation' may often threaten the validity of that approach. Furthermore, routine data do not usually include households' characteristics (e.g. wealth, gender), and so limit the capacity of the research to make detailed conclusions. In the case of user fee removal, this means for example that a sharp rise in consultations may be detected, but it is still unclear whether this is due to increased access for previously excluded groups, or just that those who already were using the facility are now attending with greater frequency. Other research tools can help investigate further why and how an intervention has had its impact. Qualitative work can prove very useful (Mills *et al*. 2008), as can small-scale household surveys that would investigate the extent to which different population groups might benefit from the observed changes.

Conclusion

Up until now most studies in low-income countries that have attempted to measure the impact of a change in user fee policy on utilization have not made a good use of longitudinal data collected in health systems. Nevertheless, this article has clearly outlined that segmented intervention analysis provides an appealing methodological framework to assess the impact of policy changes on outcomes that are routinely monitored by routine information systems. In the absence of a more robust experimental design, this approach should be favoured over a simple 'before and after' computation of summary statistics of data series. More rigorous use of existing facility data to monitor and evaluate the effects of health policy interventions would be welcome and may even lead to more attention being paid to the quality of such data and how it is collected and processed.

Acknowledgements

I thank Dr. Natasha Palmer for her guidance and useful comments on early drafts of this manuscript. I also acknowledge the Ministry of Health in Zambia, in particular M. Kaliki, who provided data from the Health Management and Information System that were used in the example presented here.

Funding

This work was undertaken while the author was funded by the UK Department for International Development (DfID).

References

Akashi H, Yamada T, Huot E, Kanal K, Sugimoto T. 2004. User fees at a public hospital in Cambodia: effects on hospital performance and provider attitudes. *Social Science & Medicine* **58**: 553–64.

Box GE, Tiao GC. 1975. Intervention analysis with application to economic and environmental problems. *Journal of American Statistical Association* **70**: 70–9.

Box GEP, Jenkins GM. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Brockwell PJ, Davis RA. 2002. *Introduction to Time Series and Forecasting*. New York: Springer.

Burnham G, Pariyo G, Galiwango E, Wabwire-Mangen F. 2004. Discontinuation of cost sharing in Uganda. *Bulletin of the World Health Organization* **82**: 187–95.

Chan SSC, Wong DCN, Fong DYT *et al.* 2009. Short-term impact of new smoke-free legislation on the utilization of a quitline in Hong Kong. *Nicotine & Tobacco Research* **11**: 356–61.

Cochrane Effective Practice and Organisation of Care Review Group. 2002. *The data collection checklist*. Online at: http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf, accessed 5 May 2009.

Collins D, Quick J, Musau S, Kraushaar D, Hussein I. 1996. The fall and rise of cost sharing in Kenya: the impact of phased implementation. *Health Policy and Planning* **11**: 52–63.

Cook T, Campbell D. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally College Publishing Company.

Duflo E, Glennerster R, Kremer M. 2004. Randomized evaluations of interventions in social science delivery. *Development Outreach*. March. Online at: http://www1.worldbank.org/devoutreach/march04/.

Fitzmaurice GM, Laird NM, Ware JH. 2004. *Applied Longitudinal Analysis*. New York: Wiley.

Grimshaw J, Alderson P, Bero L *et al.* 2003. Study designs accepted for inclusion in EPOC reviews. *EPOC Newsletter*, March 2003. Cochrane Effective Practice and Organisation of Care Review Group. Online at: http://epoc.cochrane.org/newsletters.

Hamilton JD. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Ho AKF, Wan ATK. 2002. Testing for covariance stationarity of stock returns in the presence of structural breaks: an intervention analysis. *Applied Economics Letters* **9**: 441–7.

Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee TC (eds). 1985. *The Theory and Practice of Econometrics*. New York: John Wiley & Sons.

Kipp W, Kamugisha J, Jacobs P, Burnham G, Rubaale T. 2001. User fees, health staff incentives, and service utilization in Kabarole District, Uganda. *Bulletin of the World Health Organization* **79**: 1032–7.

Lagarde M, Palmer N. 2008. The impact of user fees on utilisation of health services in low and middle income countries – how strong is the evidence? *Bulletin of the World Health Organization* **86**: 839–48.

Mbugua JK, Bloom GH, Segall MM. 1995. Impact of user charges on vulnerable groups: the case of Kibwezi in rural Kenya. *Social Science & Medicine* **41**: 829–35.

Mcdowall D, McCleary R, Meidinger EE, Hay RA. 1980. *Interrupted Time Series Analysis*. Newbury Park, CA: Sage Publications.

Meuwissen LE. 2002. Problems of cost recovery implementation in district health care: a case study from Niger. *Health Policy and Planning* **17**: 304–13.

Mills A, Gilson L, Palmer N, Lagarde M. 2008. What do we mean by rigorous health-systems research? *The Lancet* **372**: 1527–9.

Mohr LB. 1995. *Impact Analysis for Program Evaluation*. Thousand Oaks, CA: Sage.

Moses S, Manji F, Bradley JE *et al.* 1992. Impact of user fees on attendance at a referral centre for sexually transmitted diseases in Kenya. *The Lancet* **340**: 463–6.

Nabyonga J, Desmet M, Karamagi H *et al.* 2005. Abolition of cost-sharing is pro-poor: evidence from Uganda. *Health Policy and Planning* **20**: 100–8.

Ramsay C, Grilli R, Grimshaw J. 2001. Robust methods for reanalysis of interrupted time series designs for inclusion in systematic reviews. 9th International Cochrane Colloquium, Lyon, France, 9–13 October 2001.

Ranson MK, Sinha T, Morris SS, Mills AJ. 2006. CRTs—cluster randomized trials or ''courting real troubles'': challenges of running a CRT in rural Gujarat, India. *Canadian Journal of Public Health* **97**: 72–5.

Ridde V. 2003. Fees-for-services, cost recovery, and equity in a district of Burkina Faso operating the Bamako Initiative. *Bulletin of the World Health Organization* **81**: 532–8.

Savedoff WD, Levine R, Birdsall N. 2006. When will we ever learn? Improving lives through impact evaluation, Washington, DC: Evaluation Gap Working Group, Center for Global Development.

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.

Singer JD, Willett JB. 2002. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Snijders TAB, Bosker RJ. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Van Driel ML, Vander Stichele R, Elseviers M *et al.* 2008. Effects of an evidence report and policies lifting reimbursement restrictions for acid suppressants: analysis of the Belgian national database. *Pharmacoepidemiology and Drug Safety* **17**: 1113–22.

Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. 2002. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy & Therapeutics* **27**: 299–309.

Wilkinson D, Gouws E, Sach M, Karim SS. 2001. Effect of removing user fees on attendance for curative and preventive primary health care services in rural South Africa. *Bulletin of the World Health Organization* **79**: 665–71.

Zhang Y, Adams AS, Ross-Degnan D, Zhang F, Soumerai SB. 2009. Effects of prior authorization on medication discontinuation among medicaid beneficiaries with bipolar disorder. *Psychiatric Services* **60**: 520–7.