ORIGINAL RESEARCH ARTICLE

Appl Health Econ Health Policy 2005; 4 (2): 87-98
1175-5652/05/0002-0087/$34.95/0

What is Required to Evaluate the Impact of Pharmaceutical Reference Pricing?

Jaume Puig-Junoy

Department of Economics and Business, Research Centre for Health and Economics (CRES), Pompeu Fabra University, Barcelona, Spain

Abstract

Objective: To describe empirical studies evaluating the impact of reference pricing (RP) interventions in pharmaceutical markets in order to discuss the requirements for these evaluations.

Methods: Ten studies were included in this review. For each study, the nature of the intervention, the nature of the data available, the nature of the question to be answered and the requirements of the evaluation method were examined through a questionnaire. The most frequently used evaluation method was the conventional before-after estimator, and only three studies used the difference-in-differences method.

Results: Nine studies evaluated a therapeutic RP system and one evaluated a generic RP system. All of the papers reported how the reference price level was established, but only one study directly reported the updating frequency and criteria of the RP system. In four studies, details of simultaneous interventions were not reported. There is no paper providing evidence on overall social welfare impact. Four papers estimated the impact of intervention on the consumer price of drugs covered by the RP system. Only one provided information about the impact on the price of related drugs not covered by the system. Three studies included an outcome variable for the use of health services. The impact of RP intervention on the level of competition in the market for those medicines covered by the system was reported in only three of ten papers.

Discussion/conclusion: Despite the rigorous effort made to evaluate the impact of RP policies in some countries, several limitations may affect both their internal validity (the nature of the data available, statistical problems common to non-experimental data, etc.) and their external validity (heterogeneity in the nature of the intervention).

Since its explicit introduction in Germany in 1989, many public and private insurers in several countries have adopted pharmaceutical reference pricing (RP) as a strategy for cost containment. Various designs have been applied in countries such as Australia, Canada (British Columbia and Nova Scotia), Denmark, Greece, Hungary, Italy, Lithuania, The Netherlands, New Zealand, Norway, Spain, the Czech Republic and Sweden. Despite its abandonment in Norway in 2001,[1] a common factor leading to the increased diffusion of RP policies has been concern about the increasing relative impact of pharmaceutical expenditure on public funds, and the effectiveness of RP in curbing expenditure rise. An RP policy has also been proposed for a Medicare drug benefit in the US.[2]

The basic features of an RP system are simple.[3-5] The third-party payer exercises buying power by setting an identical maximum reimbursement price to be paid for a group of pharmaceutical products ('clusters'). The insured patient pays the difference if the chosen medicine is more expensive. This co-payment may be avoided if the drug does not exceed the reference price. By limiting the level of reimbursement, RP aims to reduce the price of referenced products either through a relative decrease in the demand for high-priced products (a demand-side approach) or through cuts in drug prices by encouraging self-restraint (a supply-side approach). In this sense, it is a policy aimed at fostering price competition through prices by channelling public funding towards the lowest-priced products.

However, the RP policy systems employed by each insurer differ greatly in detail and scope. In fact, we should think of RP as a family of many different pharmaceutical insurance coverage policies. 'Clusters' are defined in terms of their interchangeability.

This may be interpreted strictly or loosely depending on whether the criterion used is chemical equivalence (generic RP) or pharmacological or therapeutic equivalence (therapeutic RP).

There is a wide range of RP policies. They vary according to:

- equivalence level and criteria;
- determination of the reference price level;
- inclusion of patented drugs;
- therapeutic groups included;
- system of exemptions from the co-payment associated with RP;
- level and type of pre-existing co-payment;
- incentives for doctors and pharmacists;
- price regulation system;
- number of producers competing in the market;
- possibilities of parallel trade;
- relationship between domestic prices and price regulation in other countries.

This heterogeneity means that the results of evaluations of different RP interventions are difficult to compare, as they effectively correspond to different policies.

The social welfare impact of a policy such as an RP system, which involves changes in the pharmaceutical insurance coverage, depends on how patients, pharmaceutical firms, health service providers (doctors, hospitals, pharmacists, etc.) and the insurers themselves respond. Evaluation of this impact is essential for determining the effectiveness of such a policy, and also for establishing the advantages and disadvantages of different types of RP systems.

To date, a number of works have highlighted significant limitations to empirical research on RP.[1] These hinder precise determination of the effective impact of this type of policy. In a review of the literature published up to 1998, López-Casasnovas and Puig-Junoy[3] stressed two major limitations of studies on RP.

First, most of the literature was purely descriptive, which made it difficult to draw causal inferences on effective impact. However, this situation has changed substantially in recent years with the publication of many empirical studies using rigorous methods and individual patient-level data.

Second, in studies with highly aggregated data, it is very difficult to identify precisely the effects of RP. It is hard to separate these effects from those related to the occurrence of other coincident policy changes, and indeed other factors and trends that influence the behaviour of pharmaceutical prescriptions.

This article reviews empirical studies that evaluate the impact of RP in order to discuss the requirements of such studies. The nature of the intervention, the nature of the data available, the nature of the question to be answered and the requirements of the evaluation method were examined by questionnaire.

Reference Pricing (RP) Evaluation Framework

The evaluation problem consists of the measurement of the impact of the RP policies on a set of well defined outcome variables in an identified group of individuals.[6] In order for the evaluation of the RP policy to be complete, it should be possible to establish the impact exerted on each of the expected effects, as regards both use of resources and health state (outcome variables). It should also clearly indicate the situation with which the results of the application of this intervention are being compared (counterfactual).

Outcome Variables

The choice of outcome variables will determine whether it constitutes a partial or a full impact evaluation. For the latter, the outcome variables must reflect the effect on drug prices, consumption of medicines covered and not covered by the RP system, use of other health service inputs (other health services, time spent by doctors and pharmacists on counselling patients, and management and administration resources), health state and other observed health-related effects, in addition to including overall costs and changes in cost distribution.[7] The impact on overall costs may be represented by changes in prices, consumption and input use (consumption effect), and changes in the distribution of costs between the different agents in the market. Outcome variables of interest may be different according to the type of RP policy to be evaluated (i.e. generic or therapeutic RP). For example, if benefit and risk are different for the drugs included in the same 'cluster' (therapeutic RP), then outcome variables representing healthcare use and health state, and their changes according to patient income levels, become more important.

The price effect can manifest itself not only in a variation in the price of those pharmaceuticals that fall within the RP system (pure price effect), but also in a possible variation in the price of pharmaceuticals that are not subject to RP (indirect price effect). When producers are forced to make sizeable cuts in the prices of products covered by the RP system, they may attempt to recover some of the lost revenue by increasing the sales of non-covered products and by introducing new products at high prices.

RP policies not only affect the price of pharmaceuticals, they can also affect the use (i.e. amount) of pharmaceuticals and other health services. The effect depends on the elasticity of demand with respect to changes in the co-payment associated with RP. It should be noted that prescribed volumes are usually converted into defined daily doses to reflect intensity of use more accurately.

The consumption of pharmaceuticals not included in the RP system can also change with the introduction of RP. This substitution effect on consumption may reflect decisions taken by the

patient, the prescriber and/or the influence exerted by the pharmaceutical industry on prescription decisions. The use of other resources may also be affected, including health service visits, time spent by health professionals and pharmacists with patients, and those assigned to the management and administration of the RP system (e.g. exemption management, that is, additional time required to handle exemption requests). The substitution effect may affect the health of the patient and his/her use of health services. Also, pharmacists need extra time to explain reimbursement rules and substitution to the patient.

Non-Observed Counterfactual

The ideal evaluation of an RP system should compare the results of the application of this intervention with what would have happened without the intervention. In reality, this cannot be achieved. The next best thing is to compare a group of individuals who took part in the intervention (a treatment group) with a different group who did not participate in the intervention (a control group). If the outcome variable could be observed for individuals in each group, and the groups are otherwise identical, the central counterfactual problem can be overcome. However, this would require performing a random experiment along the lines of randomised clinical trials, which is not feasible in most cases. In most empirical evaluations there is no control group available. Hence, when evaluating the impact of RP, we are faced with the problem that the counterfactual is not observed and has to be estimated using quantitative methods.[8] RP effects are usually assessed in retrospect, using service statistics that have been routinely collected for other purposes.

The counterfactual is very important in the evaluation of an intervention. It is the scenario with which the post-intervention situation is compared. The standard approach is to estimate a counterfactual equivalent to the situation that would have occurred in the absence of the intervention, and without applying any other type of intervention (i.e. the status quo). However, it should be stressed that this is not the only counterfactual of interest. We may wish to compare the impact of the intervention with the impact from a different type of RP intervention (e.g. an aggressive policy designed to encourage competition by means of incentives for substitution with the lowest-cost bioequivalent drug by pharmacists) or with a variant of the same type of intervention (such as an RP system in which the reference level is always equal to the price of the lowest-cost drug in the same group, as opposed to the average of the observed prices). In these cases, the objective would be to compare the impact of the current intervention with the alternatives that were not adopted in practice. The counterfactual would then have to be constructed by simulating the effects of these other interventions.

Evaluation Methods

It is impossible to infer causal relationships regarding the impact of RP using studies with highly aggregated data and in the absence of some type of statistical design to attempt to contrast hypotheses about the magnitude and significance of the impact. A large proportion of the RP literature[3,4] employs simple before-after comparisons for the health system. Therefore, it can only provide insight into general trends, without controlling for the influences of other policies or tendencies inherent in the market.

Experimental studies are uncommon in economics, and obtaining experimental data, even when feasible, can be expensive.[9] There are several empirical policy impact evaluation methods used to construct the counterfactual. The most widely used methods with non-experimental data include the cross-section estimator, the before-after estimator, the difference-in-differences estimator, the matching method, the selection model and the structural simulation model.[6,10]

Methods

Selection of Empirical Papers

The list of papers in a previous review of the literature on RP[3] was used as the starting point for this study. A complementary, systematic bibliographical search (of papers published in English up to August 2002) was performed on the MEDLINE and EconLit databases. The search strategy required 'reference price' or 'reference pricing' and at least one of 'pharmaceutical(s)' or 'medicine(s)' or 'drug(s)' in the title, abstract or medical subject heading (MeSH®). Additionally, a manual search was carried out to identify unpublished academic working papers and book chapters in English.

Impact evaluation studies of RP policies using non-experimental data were selected for inclusion in this study. Hence, this study reports on only impact evaluation papers relating to RP systems introduced effectively by an insurer and that test hypotheses on well defined outcome measures and construct an explicit or implicit counterfactual. In the event of duplicate publication, the most comprehensive report was chosen.

The following five types of paper were excluded.

1. Those that did not report the empirical results of an RP impact evaluation, but addressed descriptive, methodological or theoretical issues that could be of relevance to pharmaceutical reimbursement.

2. Papers that used descriptive aggregate data at a national or regional level, simply comparing observed data before and after RP introduction. These papers did not construct the unobserved counterfactual for the after period if the intervention had not occurred, nor did they take into account likely confounding effects (e.g. with control for confounding effects by adding covariates to regression models). Furthermore, they did not justify the assumption that, if the intervention had not taken place, the situation in the after period for the outcome variables would have been the same as in the before period.[11-14] Also, papers that did not even attempt to fit a trend line to the pre-policy period were excluded.

3. Papers presenting an untested hypothesis and/or without data to support the conclusions, and also empirical studies without (or with insufficient) presentation of the statistical method employed to obtain estimated impacts.

4. Papers presenting individual case studies without any defined statistical method to test hypotheses on the impact of the RP intervention.[15]

5. Empirical papers providing a simulation for the future effect of RP, but without evaluating the impact of this policy in a real RP context.[16,17]

Review Process

The criteria for judging the quality of the RP evaluation method were derived specifically for this study from mainstream economic and econometric texts on policy evaluation,[6,10] and from surveys on international experience with RP.[3-5,18] The aspects taken into account for evaluating the quality of the policy impact evaluation method, and for organising the questionnaire, were as follows: the nature of the intervention; the nature of the data available; the nature of the question to be answered; and the requirements of the evaluation method. Twenty questions were developed covering these aspects. These are listed and explained in the following sections.

The Nature of the Intervention

Question 1: "Did the study report the equivalence criteria used in the RP intervention to group medicines?"

Question 2: "If relevant for the outcomes measured, did the study report other relevant details of the RP intervention: (i) the reference price level; (ii) the updating frequency and criteria; and (iii) the exemption system?"

Question 3: "If relevant for the outcomes measured, did the study report details of the price regulation and/or the prevailing reimbursement systems?"

Given the heterogeneity of RP policies, the review sought to identify the main characteristics of the RP intervention. This is very important if the results of the evaluation are to be used to predict the likely impact of RP policies in other countries or populations. The regulatory framework in which RP has been applied is important for the impact of this policy in many aspects. For example, price changes after the introduction of RP may be constrained by the existing price regulation system, and so the observed effect may not be universally applicable.

Question 4: "Did the study state the absence of other simultaneous interventions during the observation period, or report details of simultaneous interventions?"

The confounding effects of other regulatory policies or influential factors have been mentioned repeatedly as a major limitation of RP literature. Schneeweiss et al.[19] stated that their overall conclusions about the impact of RP in Germany were biased because of confounding factors such as the introduction of drug budgets, and the many changes due to the unification of Germany.

The Nature of the Data Available

Question 5: "Does the dataset contain information at individual patient level before and after the introduction of the RP intervention?"

Question 6: "If relevant, does the dataset contain information for individual medicine prices before and after the introduction of the RP intervention?"

Time-series comparisons of before and after data for the whole population are not the only way of evaluating RP impact. It is also possible to group patients according to the impact of RP in terms of selected outcomes. For example, we could consider those patients who switch from a higher priced product to a reference drug (the treatment group) and those who stay on a high-priced drug by getting a policy exemption or by paying the difference above the RP level (the control or comparison group).[20]

Question 7: "If the answer to question 5 and/or question 6 was 'yes', does the dataset contain cross-sectional information for a treatment group and a comparison group for the same time period?"

Question 8: "Does the period after the introduction of RP cover more than 1 year?"

Question 9: "Does the dataset contain information on price or consumption of pharmaceuticals aggregated for a period not exceeding 1 month?"

The Nature of the Question to Be Answered

Question 10: "Were the following groups of outcome variables estimated and reported in the results of the study: (i) price of drugs under RP (pure price effect); (ii) price of drugs not covered by RP (indirect price effect); (iii) drug use; (iv) health services use; (v) prescriber and dispenser time; (vi) health state; (vii) insurer and patient drug expenditures; and (viii) industrial research and innovation (research and development)?"

Question 11: "Did the study report any empirical evidence on the impact on consumer price competition for medicines covered by the RP system?"

The outcome variables determine whether the study can present a partial or a full impact evaluation. Although RP policies are designed to promote price competition, it has been hypothesised that the granting of substitution authorisation to pharmacists could result in competitive discounting to pharmacists and fail to benefit payers and patients. This review is intended to examine not only the degree of completeness of the outcome variables (question 10), but also likely changes in consumer price competition.

The Requirements of the Evaluation Method

Question 12: "Did the study justify the absence of influence of unobserved variables on the outcome variables?"

Question 13: "Did the study report evidence on the absence of an anticipation effect just before the RP intervention took place, or were methods given to evaluate their influence on the outcome variables?"

Anticipation effects may greatly bias RP impact evaluation. For example, anticipation of the introduction of RP may lead to an early reduction in the price of the pharmaceutical, bringing it close to the reference level. Alternatively, anticipatory prescribing or 'stockpiling' by patients may occur, in order to avoid a possible co-payment after the introduction of RP.

Question 14: "Did the study justify the choice of the before and after periods?"

Question 15: "If the answer to question 5 and/or question 6 was 'yes', was the before period equal to or longer than 1 year and not shorter than the after period?"

The use of a before-after estimator requires a long period before the intervention. Extrapolations for periods longer than the base period are unlikely to be acceptable. Impact estimates of the RP intervention might be sensitive to the choice of the period for comparison.

Question 16: "If other simultaneous interventions were reported, were methods given for the estimation of the marginal impact of the RP intervention isolated from other interventions?"

Question 17: "If many periods of before-period data were used to extrapolate the counterfactual state in the after period, were extrapolative methods given that included variables other than a simple linear trend across time?"

Extrapolation using a linear time trend implies a very simple relationship. There are other factors to be considered in the extrapolation of monthly drug use and expenditure, such as seasonal (a monthly indicator variable) and individual characteristics (age, sex, previous medication use, education, income, etc.).

Question 18: "If many periods of the before-period data were used to extrapolate the counterfactual state in the after period, did the study report confidence intervals for extrapolated data or standard errors for the estimated parameters?"

Question 19: "If the answer to question 7 was 'yes', did the study justify the absence of a selection problem or were methods given to adjust for this problem?"

A selection problem would arise if the intervention group and the comparison group respond differently to the presence or absence of the intervention.

Question 20: "If the answer to question 7 was 'yes', did the study report evidence indicating that variables influencing the outcome variables but not related to the RP intervention were very similar between the treatment and comparison group?"

Results

Characteristics of the Papers

Ten RP impact evaluation studies survived the five exclusion criteria and were included in the review.[21-30] Accepted papers were concentrated among a few authors and countries. Seven studies (70%) were undertaken on British Columbia (Canada), and these involved three academic research centres: Harvard University (three), McMaster University (three) and the University of Washington (one).[31] Consequently, RP policy in British Columbia since 1995 has been one of the most rigorously and extensively evaluated pharmaceutical policy interventions. The other three studies evaluated RP intervention in Germany (two) and in Sweden (one).

The majority of the studies were reported in clinical and public health journals: *Canadian Medical Association Journal* (three), *New England Journal of Medicine* (one), *Medical Care* (one) and *Journal of the American Geriatric Society* (one). One appeared in the *Rand Journal of Economics*. The other three studies correspond to still unpublished working papers (two) and a book chapter.

The empirical literature on RP impact evaluation is very recent, with eight of ten studies being completed after 1999. The number of studies covered in this review is minute compared with the vast literature on RP policy.[3]

Results of the Review

The results of the review are summarised in table I.

Nine studies evaluated a therapeutic RP system and one evaluated a generic RP system (Sweden). From question 1, the equivalence criteria were identified. Given the heterogeneous nature of RP interventions, other details are needed for a more precise

Table I. Results of the review of the methods used in reference pricing (RP) impact evaluation studies

Question	Number 'yes' (%)	Number 'no' (%)	Denominator (%)
The nature of the intervention			
1. Did the study report the equivalence criteria used in the RP intervention to group medicines?	10 (100)	0 (0)	10 (100)
2. If relevant for the outcomes measured, did the study report other relevant details of the RP intervention:			
(i) the reference price level;	10 (10)	0 (0)	10 (100)
(ii) the updating frequency and criteria;	1 (10)	9 (90)	10 (100)
(iii) the exemption system?	5 (71.4)	2 (28.6)	7 (100)
3. If relevant for the outcomes measured, did the study report details of the price regulation and/or the prevailing reimbursement systems?	5 (100)	0 (0)	5 (100)
4. Did the study state the absence of other simultaneous interventions during the observation period stated, or report details of simultaneous interventions?	6 (60)	4 (40)	10 (100)
The nature of the data available			
5. Does the dataset contain information at individual patient level before and after the introduction of the RP intervention?	6 (60)[a]	4 (40)	10 (100)
6. If relevant, does the dataset contain information for individual medicine prices before and after the introduction of the RP intervention?	0 (0)	10 (100)	10 (100)
7. If the answer to question 5 and/or question 6 was 'yes', does the dataset contain cross-sectional information for a treatment group and a comparison group for the same time period?	3 (50)[b]	3 (50)	6 (100)
8. Does the period after the introduction of RP cover more than 1 year?	8 (80%)	2 (20)	10 (100)
9. Does the dataset contain information on price or consumption of pharmaceuticals aggregated for a period not exceeding 1 month?	7 (70)	3 (30)	10 (100)
The nature of the question to be answered			
10. Were the following groups of outcome variables estimated and reported in the results of the study:			
(i) price of drugs under RP (pure price effect);	4 (40)	6 (60)	10 (100)
(ii) price of drugs not covered by RP (indirect price effect);	1 (10)	9 (90)	10 (100)
(iii) drug use;	6 (60)[c]	4 (40)	10 (100)
(iv) health services use;	3 (30)	7 (70)	10 (100)
(v) prescriber and dispenser time;	0 (0)	10 (100)	10 (100)
(vi) health state;	1 (10)	9 (90)	10 (100)
(vii) insurer and patient drug expenditures;	4 (40)	6 (60)	10 (100)
(viii) industrial research and innovation (research and development)?	0 (0)	10 (10)	10 (100)
11. Did the study report any empirical evidence on the impact on consumer price competition for medicines covered by the RP system?	3 (30)	7 (70)	10 (100)
The requirements of the evaluation method			
12. Did the study justify the absence of influence of unobserved variables on the outcome variables?	3 (30)	7 (70)	10 (100)
13. Did the study report evidence on the absence of an anticipation effect just before the RP intervention took place, or were methods given to evaluate their influence on the outcome variables?	4 (40)	6 (60)	10 (100)

Continued next page

Table I. Contd

Question	Number 'yes' (%)	Number 'no' (%)	Denominator (%)
14. Did the study justify the choice of the before and after periods?	1 (10)	9 (90)	10 (100)
15. If the answer to question 5 and/or question 6 was 'yes', was the before period equal to or longer than 1 year and not shorter than the after period?	4 (40)	6 (60)	10 (100)
16. If other simultaneous interventions were reported, were methods given for the estimation of the marginal impact of the RP intervention isolated from other interventions?	4 (80)	1 (20)	5 (100)
17. If many periods of before-period data were used to extrapolate the counterfactual state in the after period, were extrapolative methods given that included variables other than a simple linear trend across time?	8 (80)	2 (20)	10 (100)
18. If many periods of the before-period data were used to extrapolate the counterfactual state in the after period, did the study report confidence intervals for extrapolated data or standard errors for the estimated parameters?	10 (100)	0 (0)	10 (100)
19. If the answer to question 7 was 'yes', did the study justify the absence of a selection problem or were methods given to adjust for this problem?	2 (67)	1 (33)	3 (100)
20. If the answer to question 7 was 'yes', did the study report evidence indicating that variables influencing the outcome variables but not related to the RP intervention were very similar between the treatment and comparison group?	2 (67)	1 (33)	3 (100)

a There were patient-level data for other morbidity and mortality variables in two studies.

b One study only included information on a treatment group and a comparison group for morbidity outcome variables. One study considered non-coincident time periods for both periods.

c One study estimated the substitution effect in drug consumption.

definition of the type of intervention (questions 2 and 3). All the papers reported how the reference price level was established, but only one study directly reported the updating frequency and criteria of the RP system. Exemption systems may be an important feature of therapeutic RP systems. However, in two studies the information about the rules used to obtain exemptions from the RP system were not reported, despite being relevant for the outcomes measured in the study.

Often in pharmaceutical markets, several interventions occur at the same time. In four studies, details of simultaneous interventions were not reported, and there was no statement confirming the absence of such interventions during the observation period (question 4).

Six studies used patient-level datasets, although only four of them contained detailed information on pharmaceutical consumption (question 5). The rest of the papers employed aggregate measures of consumption for a given population. All the patient-level datasets were for studies on British Columbia, Canada. However, none of these studies was able to use datasets with individual prices for each product, and all of them employed average price measures for the products in a group or with the same active ingredient and/or form of presentation (question 6).

Among these six studies, only three analysed the behaviour of the outcome variables for two different population groups, a treatment group and a comparison or control group, for the same period of time (question 7). The rest of the papers used a before and after comparison. Two papers adopted a very short-run perspective for the impact evaluation, using datasets containing information covering <1 year from the introduction of RP (question 8). In most cases, information databases contained aggregate information for monthly periods, although three studies used aggregate information for longer periods (question 9).

There is no paper providing evidence on overall social welfare impact (question 10). Also, the impact of RP interventions on prescriber and dispenser time and the impact on industrial research and development (dynamic efficiency) were not evaluated in any of the reviewed studies. Six of ten studies included information on the impact of the RP system on consumption of included medicines (prescriptions, defined daily doses, probability of stopping therapy, etc. per person). However, authors rarely made a distinction among short-, medium- or long-term outcomes. This is the most frequent type of outcome variable. Only one of the studies was also able to report information on the impact of RP on other related medicines not currently under the RP system (the substitution effect in drug consumption for nitrates and anti-anginals).

Four papers estimated the impact of the intervention on the consumer price of drugs covered by the RP system (the pure price effect). However, only one provided information about the impact on the price of related drugs not covered by the system (indirect

price effect) [question 10]. Also, despite numerous statements in the literature about the separate impact of RP on insurer cost and expenditures, in this review only four papers were found that provided estimates of these impacts. In some papers, although not explicitly stated, the evaluation of the impact on expenditures was restricted to the impact on insurer cost.

Three studies included an outcome variable for the use of health services (physician visits, hospitalisations, length of stay, emergency room visits, admissions to long-term care facilities, likelihood of bypass surgery, rate of revascularisations, etc.). These studies were all evaluating therapeutic RP interventions in British Columbia, and the variables were used, sometimes in an explicit form, as proxies of health state. The authors of one study explicitly stated that they were "using change in healthcare utilisation as a proxy for changes in health status".[23] In fact, a health state variable such as mortality appeared in only one study (cardiovascular-related mortality). Two papers included indirect health state indicators, such as the use of what may be considered to be rescue drugs for patients with severe symptoms. Health-related quality of life was not considered as an outcome variable in any of the reviewed studies.

The impact of the RP intervention on the level of competition in the market for those medicines covered by the system was reported in only three of ten papers (question 11). These papers dealt with the Dutch and Swedish RP systems. The impact of RP on drug prices depends crucially on resulting changes in price competition. Therefore, it is important to extrapolate price competition trends for the after period in the absence of RP, in order to construct the counterfactual.

The most frequently used evaluation method was the conventional before-after estimator, and only three studies used difference-in-differences. One study employed both methods with the same dataset. In only three papers did the authors justify the absence of influence of unobserved variables on the outcome variables (question 12). Although an anticipation effect has been described in the literature (anticipatory prescribing or 'stockpiling'), in six of ten papers its effects were not explicitly reported, nor were methods given to evaluate their influence on the results (question 13).

All the papers used before and after periods for the same or different populations, yet the authors rarely provided justification for the choice of periods and their likely influence on the results (question 14). Six papers used data for the before period to make comparisons or extrapolations with the after period, the before RP period being ≤1 year and shorter than the after period (question 15).

Simultaneous interventions were reported for the observation period in five studies (question 16). Four of these used methods to isolate the marginal contribution of RP to changes in outcome variables.

All the studies used methods to extrapolate the counterfactual state in the after period that allowed the estimation of confidence intervals for the predicted outcome variables and hypothesis testing (question 18). Of these, only two studies employed a linear cross-time trend (question 17).

In three papers the authors used two cross-sections of population data for both periods (before and after) in a difference-in-differences approach. Two of these papers reported the possibility of selection bias and provided methods to mitigate this problem, but the third failed to justify the absence of a selection problem (question 18). Also, one of these papers did not provide evidence that the other influences on the outcome variables were very similar between the two groups (question 20). When the differences-in-differences method was used, the studies did not show that both treatment and control groups reacted in the same way to common past interventions.

Discussion

Until very recently, the bulk of RP literature remained mainly descriptive and, when empirical information was provided, only presented time trends in aggregate data. In fact, some health impact evaluations that are often cited in the literature as evidence of the health impact of RP policies were excluded by our selection criteria.[31] However, as can be observed in this review, the situation has improved and, despite the restrictive selection criteria applied, ten rigorous impact evaluation studies for RP interventions were identified.

RP systems have been applied in many countries and with different intervention designs, but, for most of these, impact evaluation studies are still not available. It is noteworthy that seven of the ten papers included in this review evaluate the reference-based pricing system operating in British Columbia.[32] This evaluation work has been possible given the availability of patient-level medical and administrative databases in British Columbia, and the participation of researchers in policy implementation and evaluation committees. This high concentration of studies means that the evidence mainly corresponds to a very specific RP design (therapeutic equivalence for some particular groups of medicines, reference level equal to the lowest-cost medicine in the group, generous and widely used exemption system, previous limits to the lowest-priced drug when there were alternatives available, generic substitution, etc.).[3] Consequently, this evidence cannot be easily transferred to other intervention designs in other countries. In contrast, little evidence is available on the

impact of RP systems specifically employed in other countries, and this is especially true for widely used generic RP systems.

The nature of the evaluation could be different depending on whether the RP interventions are generic or therapeutic. In therapeutic RP policies, the main interest has been in providing evidence of changes in patients' health after the introduction of this system. However, in generic RP systems that cluster drugs with the same active ingredient, this outcome is less relevant, and the impact on prices and expenditure remains the main factor of interest for evaluation purposes.

In both generic and therapeutic RP systems, both intended and unintended consequences of the policy should be considered in order to obtain an overall impact measure. A limitation of many studies is that they restrict their attention to some partial impacts, neglecting substitution effects with other drugs not covered by the RP system or other health services. In fact, only one study provided estimates of likely compensatory increases in expenditures on other drugs. This aspect is not always clearly stated. The implication is that conclusions about social welfare changes cannot be reached using only the perspective of the insuring or financing agency. This is because savings could have occurred at the expense of other patient costs or health-related outcomes. Also, overall conclusions on insurer costs cannot be obtained simply by estimating pure price effects, as many insurers have done in the past.

Generic RP Systems

Despite the apparent simplicity of generic RP systems where equivalence is restricted to the same active ingredient, there is strikingly little empirical evidence on their impact. One of the main arguments used in favour of the introduction of this RP system has been that of fostering price competition. However, there are three studies in our review[21,22,30] that provide evidence that price competition may even be discouraged under generic RP. This is a crucial aspect to be tested, especially because generic RP is considered an optimal insurance policy from a theoretical point of view, and because its pseudo-simplicity makes it a candidate for widespread dissemination internationally, including in less developed countries. More evidence is needed about the effects of generic RP on the entry of new generic producers into the market, and on the industry's incentives to price below the reference price.[30] These latter factors can vary according to the incentive system and the substitution capacity of pharmacists.

The effect of RP on price competition (payer-driven competition) was not an outcome of interest in most of the papers. However, this could be crucial for evaluating and providing policy implications for generic RP systems such as in The Netherlands and Spain. There, evidence indicates the absence or reduction of consumer price competition after RP, and price competition has taken the form of competitive discounting to pharmacies. The problems that remain for generic RP are how to arrive at the best design to foster price competition, and how to transfer the effects of this to patients and insurers rather than increasing the revenue of pharmacies. None of the studies analysed the impact of pharmacy incentives on RP outcomes.

When drugs are not actually interchangeable in terms of benefit and risk, health-related impacts of therapeutic RP may occur as a result of switching from a high-priced medication to a medication with a price not exceeding the reference level. However, not only the switcher's health may be compromised. Some patients may remain on the same drug as was prescribed before the introduction of the RP system. Some of these may stop treatment, reduce the dose or delay prescription renewals by reducing daily doses to postpone co-payments ('stretching').[27] Most of the authors focused their attention on health outcomes for switchers, and paid less attention to likely impacts on non-switchers.

Health-Related Impact

Sometimes the only purpose of a study is to evaluate the health-related effects of switching from one drug to another. Evidence may be obtainable through randomised trials and well designed economic evaluations using precise quantity and quality-of-life measures. Authors choosing instead to use observational methods should justify their choice.

One of the main problems in the RP evaluation literature is that most studies providing evidence on health outcomes employ health services use as a proxy for health state. Only Grootendorst et al.[24] reported changes in cardiovascular-related mortality, and none of the reviewed papers was able to provide evidence on quality-of-life changes attributable to switching from one medicine to another. Quality-of-life effects cannot be captured by such outcome measures, and long-term health effects are, in practice, difficult to measure and to attribute to switching medication. A related, but not less important, limitation observed by Schneeweiss et al.[29] is that, "the administrative databases could not provide enough detailed information to distinguish the clinical appropriateness of stopping antihypertensive medication or reducing the dose in individual cases".

Before and After Comparisons

The most common evaluation method, used in eight of these papers, was the before-after estimator. The important conditions or requirements to obtain reliable estimations with this method are: (i) between the before and after periods there should be no factor

that might affect the outcome variable other than the intervention itself (temporal homogeneity); (ii) the intervention should generate no anticipation effect; and (iii) the specific timing of the data gathering before and after the intervention should not affect the results.

Approximation errors occur when the difference between the observed value for the outcome variable in the period before the intervention and the value that would have been obtained in the absence of the intervention in the after period is non-null. Only if this difference is null can the before-after method be used in simple form. With the non-fulfilment of any of the three requirements already mentioned, this condition is effectively violated. Two different approaches can be used to solve this problem.[10] One alternative involves using the difference-in-differences method. A second, controversial, alternative is the generalisation of the before-after method by using a long time series before the intervention and extrapolating the counterfactual on this basis for the after period. This is the option employed in most RP impact evaluations. The hypothesis used is that the approximation error can be reduced to zero by extrapolating the values of the outcome variable for the after period. This extrapolation can be done, for example, by assuming that the outcome variable is a function of time and a given number of independent variables.[32]

Questions 12–18 provided examples of notable limitations (unobserved variables, anticipation effect, period choice, simultaneous interventions) in the use of the before-after approach. For example, only Marshall et al.[26] reported that "estimates of the savings attributable to the introduction of reference-based pricing … were particularly sensitive to the choice of period for comparison" and that "a sensitivity analysis was undertaken …".

The aim of the difference-in-differences method is to obtain the difference between the sample differences of the outcome variable before and after the reform for the group of participants and the group of non-participants.[10] The method can be applied when before and after data are available on individuals who participated in the intervention and those who did not. It was used in three of the studies reviewed. These compared the results of the group of patients who altered their pharmaceutical treatment as a result of the application of therapeutic RP with those of the group who maintained it unchanged.

The main advantage of this method is that it eliminates any bias from an outside influence, provided this influence is identical in the two groups.[9] For this reason, it is important that the participating group be as similar as possible to the non-participating group. The underlying hypothesis is that, in the absence of the intervention, the mean variation between the before and after periods would be the same for the two groups.

Some of the limitations in the use of the approach in RP impact evaluation have been summarised by Grootendorst et al.[24] The evaluation has to identify the contribution of RP to switching probabilities, because not all switches can be attributed to RP. This approach suffers from selection bias as, for example, switchers might be less healthy than non-switchers.

The before-after method is used in an attempt to identify intervention-related changes between the two periods affecting both the level and the time trend of the outcome variable. This estimation of the impact of the intervention is intended as a measure of the average gain of the intervention (average change in the outcome variable caused by exposing an individual chosen at random to the policy in question). Schneeweiss et al.[19] criticise this measure. Although it is useful for policy decision makers, it is an average of the positive and negative effects that may be experienced by some individuals (aggregate net effect). Also, certain undesired effects of the intervention may affect a small number of individuals (e.g. those who change their pharmacological treatment as a result of RP). This may not be detected by statistical tests on aggregate results.

The criticisms voiced by Schneeweiss et al.[20] indicate the advisability of distinguishing between the average outcome of the participants obtained with the difference-in-differences method and the average outcome of the reform or policy. This distinction is especially relevant when measuring impact on outcome variables such as health state or the use of other health services. These are only likely to be affected by the intervention in the case of those individuals who are forced to change a medicine in order to avoid the co-payment associated with RP.

Conclusions

In this review, ten impact evaluation studies survived the strict exclusion criteria. This is a very small number in comparison with the vast RP literature. However, the evidence they provide can be regarded as high quality and relatively abundant, considering that RP systems have been implemented quite recently, and in a limited number of countries. For example, although the application of conventional co-payments is far more widespread, and has a far longer history than the RP system, a recent review found only 59 articles fulfilling much less strict inclusion criteria than have been used in this study.[33]

An important policy implication arises from the preceding observations. Conclusions on the impact of RP in a specific country cannot be easily extrapolated to other countries because of heterogeneity of RP policy and pharmaceutical environment. Each way of fixing the reference price level amounts to a different policy. Also, the impacts of interventions may depend on aspects

such as the previous degree of development of the generic market. Hence, any use of evidence from one country in another must be done with caution.

None of the evaluations included in this study can be regarded as complete from a social point of view. While the selected outcome variables depend on the perspective of the evaluation, the predominant perspective taken is that of the insuring or financing agents. Even then, the reviewed results are only a partial evaluation of this perspective. There is a striking lack of attention paid to the evaluation of generic RP systems. In particular, there is clear neglect of the following: the measurement of the impact on price competition between generic producers; possible effects on the price of products not submitted to the RP system (indirect price effect); and the effects of the alternative ways of fixing the reference price level (lowest-price drug, lowest-price drug plus x%, median price, average dosage unit price of two lowest-cost products in a group, statistically derived average price of drugs in a category, etc.).

The main problem with using non-experimental data to evaluate policy impact is how to establish a credible basis for comparison. Aggregate statistics are usually inappropriate for analysing many kinds of microeconomic behaviour. Administrative databases seldom provide any information about what the participants' experiences would have been had they been enrolled on a different programme.[8] Common sources of statistical bias in studies such as these include sample selection, simultaneous interventions, the anticipation effect, omission of confounding effects, and heterogeneity in comparison groups in time-series or cross-sectional data. Evidence of these problems is provided in the Results and Discussion sections. In some cases, the use of sophisticated econometric methods made it possible to avoid or alleviate some of these problems, but in many other studies they remain unaddressed or unsolved.

As a final observation, studies evaluating therapeutic RP systems have directed their attention to measuring the impact on health outcomes. This impact has been estimated mainly through the effect on health service use or mortality, using non-experimental data to make comparisons between switchers and non-switchers. The conclusions obtained with these data may be influenced by unobserved or uncontrolled factors such as variation in severity of illness or changes in treatment compliance. At the same time, switching from one active ingredient to another with similar therapeutic properties would appear to affect quality of life rather than mortality. However, the reviewed studies reported only mortality changes or changes in health services use, which are a poor proxy for quality of life. When the main objective of the evaluation is the health outcome of the policy, non-experimental data provide no better information than experimental data from randomised controlled trials. Results coming from randomised clinical trials comparing two or more pharmaceutical treatments and estimating incremental measures of quality-adjusted life-years may provide superior evidence on the effects of switching from one medicine to the other.

Acknowledgements

The author is grateful to M. Vera for useful comments on an earlier version of the manuscript. The author received support from an unrestricted educational grant from the BBVA Foundation, Spain (Sociales 03).

The author has provided no information on conflicts of interest directly relevant to the content of this study.

References

1. Jönsson B. Reference pricing: central economic and policy issues. In: López-Casasnovas G, Jönsson B, editors. Reference pricing and pharmaceutical policy: perspectives on economics and innovation. Barcelona: Springer, 2001
2. Huskamp HA, Rosenthal MB, Frank RG, et al. The Medicare prescription drug benefit: how will the game be played? Health Aff 2000; 19 (2): 8-23
3. López-Casasnovas G, Puig-Junoy J. Review of the literature on reference pricing. Health Policy 2000; 54: 87-123
4. Danzon P. Reference pricing: theory and evidence. In: López-Casasnovas G, Jönsson B, editors. Reference pricing and pharmaceutical policy: perspectives on economics and innovation. Barcelona: Springer, 2001
5. Ioannides LL, Ibrahim E, McNeil JJ. Reference-based pricing schemes: effect on pharmaceutical expenditure, resource utilization and health outcomes. Pharmacoeconomics 2002; 20 (9): 577-91
6. Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fisc Stud 2000; 21 (4): 427-68
7. Puig-Junoy J. Evaluación de los efectos de los precios de referencia: revisión metodológica e implicaciones para la política farmacéutica. Madrid: Documento de Trabajo de la Fundación FSIS, 2003
8. Burtless G. The case for randomized field trials in economic and policy research. J Econ Perspect 1995; 9 (2): 63-84
9. Vera-Hernández M. Evaluando intervenciones sanitarias sin experimentos. Gac Sanit 2003; 17 (3): 238-48
10. Garattini L, Giuliani G, Chiara G. Is reference pricing effective in cost containment? The German experience, 1989-1992. Pharma Pricing Review 1997; 2 (7): 131-5
11. Giuliani G, Selke G, Garattini L. The German experience in reference pricing. Health Policy 1998; 44: 73-85
12. Thomas MC, Mann DM. The impact of reference pricing on clinical lipid control. N Z Med J 1998; 111: 292-4
13. Narine L, Senathirajah M, Smith T. Evaluating reference-based pricing: initial findings and prospects. CMAJ 1999; 161: 286-8
14. Zweifel P, Crivelli L. Price regulation of drugs: lessons from Germany. J Regul Econ 1996; 10: 257-73
15. Bourgault C, Elstein E, Le Lorier J, et al. Reference-based pricing of prescription drugs: exploring the equivalence of angiotensin-converting enzyme inhibitors. CMAJ 1999; 161: 255-60
16. Atella V. Drug cost containment policies in Italy: are they effective in the long run? The case of minimum reference price. Health Policy 2000; 50: 197-218
17. Heckman JJ, Lalonde RJ, Smith JA. The economics and econometrics of active labor market programs. In: Ashenfelter A, Card D, editors. Handbook of labor economics. Vol. 3. Amsterdam: Elsevier, 1999
18. Dickson M, Redwood H. Pharmaceutical reference prices: how do they work in practice? Pharmacoeconomics 1998; 14 (5): 471-9
19. Schneeweiss S, Schöffski O, Selke GW. What is Germany's experience on reference based drug pricing and the etiology of adverse health outcomes or substitution? Health Policy 1998; 44: 253-60

20. Schneeweiss S, Maclure M, Walker AM, et al. On the evaluation of drug benefits policy changes with longitudinal claims data: the policy maker's versus the clinician's perspective. Health Policy 2001; 55: 97-109
21. Pavcnik N. Do pharmaceutical prices respond to potential patient out-of-pocket expenses? Rand J Econ 2002; 33 (3): 469-87
22. Ekelund M. Generic entry before and after the introduction of reference prices. In: Ekelund M, editor. Competition and innovation in the Swedish pharmaceutical market. Stockholm: Stockholm School of Economics, 2001
23. Grootendorst P, Dolovich LR, O'Brien BJ, et al. Impact of reference-based pricing of nitrates on the use and costs of anti-anginal drugs. CMAJ 2001; 165 (8): 1011-9
24. Grootendorst PV, Dolovich LR, Holbrook AM. The impact of reference pricing of cardiovascular drugs on health care costs and health outcomes: evidence from British Columbia. Volume II: technical report. SEDAP research paper no. 71 (McMaster University QSEP Report Series No. 370). Hamilton (ON): McMaster University, 2002
25. Hazlet TK, Blough DK. Health services utilization with reference drug pricing of histamine(2) receptor antagonists in British Columbia elderly. Med Care 2002; 40 (8): 640-9
26. Marshall JK, Grootendorst PV, O'Brien BJ, et al. Impact of reference-based pricing for histamine-2 receptor antagonists and restricted access for proton pump inhibitors in British Columbia. CMAJ 2002; 166 (13): 1655-62
27. Schneeweiss S, Maclure M, Soumerai SB. Prescription duration after drug copay changes in older people: methodological aspects. J Am Geriatr Soc 2002; 50: 521-5
28. Schneeweiss S, Walker AM, Glynn RJ, et al. Outcomes of reference pricing for angiotensin-converting enzyme inhibitors. N Engl J Med 2002; 346 (11): 822-9
29. Schneeweiss S, Soumerai SB, Glynn RJ, et al. Impact of reference-based pricing for angiotensin-converting enzyme inhibitors on drug utilization. CMAJ 2002; 166 (6): 737-45
30. Danzon P. Reference pricing: theory and empirical evidence. In: López-Casasnovas G, Jonsson B, editors. The economics of reference pricing and pharmaceutical policy. Barcelona: Springer Verlag, 2001: 86-126
31. Cassels A. Paying for what works: BC's experience with the Reference Drug Program as a model for rational policy making. Vancouver (BC): Canadian Centre for Policy Alternatives, BC Office, 2002
32. Heckman JJ, Robb R. Alternative methods for evaluating the impact of interventions. J Econom 1985; 30: 239-67
33. Maclure M, Nakagawa RS, Carleton BC. Evaluating evidence-based drug policies in British Columbia. In: Fox DM, Oxman AD, editors. Informing judgement: case studies of health policy and research in six countries. New York: Milbank Memorial Fund, 2002: 35-70

Correspondence and offprints: Dr *Jaume Puig-Junoy*, Department of Economics and Business, Research Centre for Health and Economics (CRES), Pompeu Fabra University, C Ramon Trias Fargas 25-27, Barcelona, 34-08005, Spain.
E-mail: jaume.puig@upf.edu