Unclassified **DAF/COMP/WP2(2016)5**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **11-Oct-2016**

English - Or. English

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

DAF/COMP/WP2(2016)5
Unclassified

Cancels & replaces the same document of 31 May 2016

Working Party No. 2 on Competition and Regulation

SECOND CAPACITY BUILDING WORKSHOP ON EX-POST EVALUATION OF COMPETITION AGENCIES' ACTIVITIES

-- Note by John Kwoka --

13 June 2016

This paper was prepared by Prof. John Kwoka (North-eastern University) to serve as a discussion paper under Item 7 at the 61st meeting of Working Party No. 2 on 13 June 2016.

The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

More documents related to this discussion can be found at:
www.oecd.org/daf/competition/evaluationofcompetitioninterventions.htm

Please contact Ms.Ania Thiemann if you have any questions regarding this document
Ania.THIEMANN@oecd.org

JT03402495

Complete document available on OLIS in its original format

English - Or. English

SECOND CAPACITY BUILDING WORKSHOP ON THE EX-POST EVALUATION OF COMPETITION AUTHORITIES' ACTIVITIES

OECD, PARIS
19 APRIL 2016

DISCUSSION PAPER BY JOHN KWOKA[*]

[*] John E. Kwoka. Jr., Northeastern University, 301 Lake Hall, Boston, Massachusetts 02115 – United States. Email: j.kwoka@neu.edu.

1. Introduction

1. The Second Capacity Building Workshop on the Ex-Post Evaluation of Competition Authorities' Activities was held at the OECD headquarters in Paris on 19 April 2016. The purpose of the Workshop was to underscore the importance of ex post evaluations of competition initiatives and to illustrate the diverse manner in which such evaluations are being, and can be, used. This note is intended to summarize and synthesize the events of that day and, based on those, to offer some observations and implications for such evaluations going forward.

2. The underlying premise of the Workshop is that competition agency activities, much like the activities of other public bodies and the actions of private entities, can be informed and improved by careful examination of past practice. For competition agencies, these practices include the choice of matters to investigate and pursue, methods of analysis and criteria for decisions, and the nature of any remedial actions that may be taken. While some ex post evaluations of competition policy have long existed, the lack of comprehensive assessments of their findings has resulted in inadequate understanding about the effects and effectiveness of past actions by competition agencies, and in diminished ability to use past experience to improve future actions.

3. In recent years, more such evaluations of specific actions taken by competition agencies in various jurisdictions have been conducted. In addition, there have now been some comprehensive compilations of existing literature, including efforts to synthesize their findings. This Workshop sought to highlight the usefulness of such evaluations by examining certain of those that have recently been conducted. It featured presentations, in summary form, of the ex post evaluations of three different competition agencies' activities, followed by discussion, comments and interventions from delegate representatives on each. The focus throughout was on the methods and the findings of these evaluations, with a view toward their use in future work by competition agencies.

4. This discussion note begins with a brief summary of the presentations, as well as comments and interventions, on the three ex post evaluations that were the subjects of this Workshop. That is followed by an analysis of the issues highlighted in the ex post evaluations that comprised this Workshop.

2. Overview of Cases at the Workshop

5. The workshop was led off by an overview of issues arising in the process of evaluating competition policies. This presentation was based on the work done for the European Commission's study of competition policies in the energy market[1] and represented a useful overview of the process involved in commissioning such an evaluation-the timeline, the data and other resource requirements, the framing of the question, resolution of uncertainties, and the interaction between the agency and the outside consultancy undertaking the work, among other matters.

6. From this initial discussion on, there was widespread agreement among all participants in the Workshop presenters, commenters, and delegatesBin the value of ex post evaluations. It seems clear that their usefulness in policy evaluation and improvement needed no further demonstration. The previous Workshop,[2] the newly released OECD guide on evaluating competition policy actions,[3] initiatives of the

1 EC, "Study on the economic impact of enforcement of competition policies on the functioning of the energy market," Brussels, 16 November 2015, Part I.

2 Capacity Building Workshop on the "Ex-Post Evaluation of Competition Authorities' Enforcement Decisions," OECD, Paris, April 2015.

International Competition Network, and widely-read summaries of EU and US policy retrospectives[4] appear to have already established their value to the satisfaction of all constituencies. Attention increasingly has turned to implementation issues-how best to frame questions, what data and other resources are necessary, how to actually test the effects of various policies, and how to use the results. These emphases were evident in the Workshop and reflect an underlying shift as more competition agencies embrace the concept of ex post evaluations and address the practicalities.

7. This initial discussion was followed by presentations of three ex post evaluations, each of which merits further description, followed by an overall assessment of their contributions.

2.1 The EC investigation of E.ON abuse of dominant position

8. In the 2000s, the largest four energy suppliers to the German wholesale electricity market accounted for about 70 percent of total sales. E.ON was the largest of the four, and was alleged to have exploited its market position in 2002-2007 in two ways: by deliberately withholding capacity in order to drive up wholesale price, and by favoring its own production affiliate for providing balancing services. As the EU Commission was conducting its investigation of these allegations in 2008, E.ON committed to divesting 5000 MW of capacity to resolve the allegations. The matter was then settled without a full investigation or proof of the allegations.

9. The evaluation was one part of a larger inquiry into whether EU competition enforcement actions have improved the functioning of EU electricity and gas markets.[5] The research strategy employed in this study involved examination of daily data on peak vs. off-peak prices from the European Energy Exchange, on the theory that effective divestiture should have reduced peak prices but have less affected off-peak prices. It found evidence consistent with this expectation, and concluded that the E.ON divestiture and Commission action did in fact succeed in reducing the wholesale price of electricity in Germany. It further examined whether this wholesale price reduction was passed through to retail consumers. While it found an effect at retail, the causal connection to the policy action was less clear.

2.2 Retail grocery mergers in the U.S. investigated by the Federal Trade Commission

10. The FTC has actively monitored the retail grocery sector in the U.S., challenging some mergers but allowing many others where market concentration is not especially high. This policy is rooted in the view that retailing is generally competitive due to ease of entry, a view that has played an important role in permitting many mergers in department stores and drug stores as well as supermarkets. A key question addressed in this paper is whether this view of entry and the resulting policy is well founded, that is, whether it has correctly distinguished anticompetitive vs. benign mergers in this sector.

---

3 "Reference Guide on Ex-Post Evaluation of Competition Agencies' Enforcement Decisions," OECD, April 2016.

4 "A Review of Merger Decisions in the EU: What Can We Learn from Ex Post Evaluations?" EC, Brussels, July 2015. F. Mariuzzo, P. Ormosi, and R. Havel, "What Can Merger Retrospectives Tell Us? An Assessment of EU Mergers", U. of East Anglia, March 2016. J. Kwoka, *Mergers, Merger Control, and Remedies*, MIT Press, Cambridge, 2015.

5 EC, "Study on the economic impact of enforcement of competition policies on the functioning of the energy market," op. cit., Part II. This study was undertaken for the EC by ICF Consultancy Services in association with DIW Berlin.

11. This working paper by FTC staff economists examines this question by compiling data on 26 common grocery items over the period 2004-2009 for 14 locations in which there were mergers or acquisitions of supermarkets.[6] These markets varied in their concentration, with six markets identified as having moderate or low concentration, and eight others as highly concentrated. Since the different levels of postmerger concentration did not result from organic growth of the companies, concentration was viewed as more-or-less exogenous. That in turn permitted inferences concerning the causal relationship between concentration and price that were statistically more reliable than conventional cross-sectional studies. This analysis found that prices often, but not always, rose after mergers in highly concentrated markets, but did not rise, and often fell, after mergers in unconcentrated markets.

2.3 The BAA airports market investigation by the UK Competition Commission.

12. In 2006 the UK Office of Fair Trade undertook a study of airport services provided by BAA. This led to a Competition Commission investigation that concluded in 2009 that there were adverse effects on competition in airport services due to common ownership of airports in the southeast (Heathrow, Stansted, Gatwick, and Southampton) and in Scotland (Edinburgh, Glasgow, and Aberdeen). Other concerns about market power stemmed from Heathrow=s dominance as a hub in the southeast and Aberdeen's relatively isolated location. Further issues were raised about the effectiveness of regulation already in force. The Competition Commission required divestiture of both Stansted and Gatwick to different purchasers, divestiture of either Glasgow or Edinburgh, and measures to improve performance at Aberdeen and under regulation more generally.

13. This study sought to evaluate these various actions by the Competition Commission.[7] Using a combination of desk research, interviews, and data analysis, it sought to measure the numerous ways in which airports competed: for carriers, for passengers, and in terms of quality of service, productive efficiency, and charges. It reported increases in take-offs and landings at the divested airports, with some indications these could be greater than otherwise would have occurred. Shares of passengers have also risen at Gatwick, Stansted, and Edinburgh, although at the latter the increase appears to be due to longer term trends. Evidence regarding available seats, service quality, aero charges, and operational efficiency is mixed. Overall, the study concluded that there is some evidence of improvements due to the policy actions.

2.4 Contributions of the case presentations

14. The specific ex post evaluations presented at this Workshop represented innovative applications of this technique. They addressed new policy evaluation questions and employed new methods of analysis, and in so doing, underscored the further challenges in its use. Substantively, for example, the cases included efforts to conduct ex post evaluations of competition policy measures beyond simple mergers. Merger retrospectives are often the most straightforward exercises, with the result that mergers have been examined considerably more often than cases where anticompetitive practices and abuse of dominance are the issues. Moreover, these Workshop cases include an ex post evaluation of a structural remedy-that is, divestiture-rather than a merger or other consolidation in the industry. Given that there have been few prior efforts to evaluate the effects of remedial competition policy, this effort addresses an interesting and less well researched question.[8]

6 D. Hosken, L. Olson, and L. Smith, "Do Retail Mergers Affect Competition? Evidence from Grocery Retailing", Federal Trade Commission, Washington, DC, September 2015.

7 "Evaluation of the 2009 Competition Commission's BAA Airports Market Investigation", London, April 2016. This evaluation was undertaken by ICF Consulting Services for the Competition and Markets Authority.

8 One effort to evaluate a merger remedy is D. Osinski, "Merger Remedies: A Retrospective Analysis of Pinnacle/Ameristar", presented at the International Industrial Organization Conference, Philadelphia, April 2016.

15. From a methodological standpoint, these three ex post evaluations also represent examples of work at the analytical frontier. Thus, the first generation of merger retrospectives essentially involved computing difference in differences using a well-chosen control group. The second generation of such studies, exemplified by one of the case studies presented in the Workshop, addresses certain econometric and causation issues that previously received inadequate attention but had the potential to produce misleading results. Similarly, the abuse of dominance and divestiture cases that were presented had to confront complex issues such as multiple overlapping policies, complicated geographic market questions, and difficulty in identifying a control group. These, too, are matters either avoided or not well addressed in earlier ex post evaluations in the literature.

16. In both of these respects-one substantive, the other methodologicalBthe ex post evaluations presented at this Workshop signified important advances in applying and improving the technique. These topics as well as others highlighted there will now be discussed in greater detail.

3. Evaluating Competition Agency Actions

17. It is useful to outline a conventional ex post evaluation of a common competition issue-a horizontal merger, with a focus on the effect on price. The standard methodology would compare the price of the relevant product that has an overlap in production or sales between the merging parties, before and after the merger. Putting aside for the moment issues of timing and other matters that will be discussed below, the most obvious concern with this comparison is that other factors might be responsible for part or all of the measured price difference. Cost changes, demand changes, and other events obviously can, and often do, affect price as well.

18. There are several empirical techniques that can serve to address this issue. The most fundamental would be to estimate econometrically a so-called structural model that sets out the underlying demand behavior and cost/supply forces in the industry. This requires complete specification of the demand function and cost relationship, including functional forms; data on all variables, whether they are of direct interest or not; and relatively sophisticated econometrics. Given these hurdles and also given that it is in fact not necessary to resolve all of them for the limited purposes of merger impact evaluation, structural estimation is rarely employed for this purpose.

19. A less burdensome alternative is so-called reduced form estimation. This, too, involves enumeration of all the relevant variables and the same data, but does not require specification of functional form, nor does it encounter the same econometric hurdles. There are some examples of this technique in the literature, but since it, too, poses some difficulties that can be avoided, recent experience does not generally resort to reduced form estimation, either.

20. So-called difference-in-differences (DID) constitutes a third and much simpler technique for controlling for other possible factors. DID avoids the need for full specification of behavior (as with structural models) or even full enumeration of variables (as in reduced form models). Rather, in order to control for influences on price other than the merger, it relies on the price change on other products that are comparable in all respects except for the merger. If truly comparable in terms of demand, cost, and other factors, the price experience of the other products--the "control group"--should therefore capture the influence of all of those factors. Taking the difference in these two differences then reveals the effect due to the merger-the only factor that differs between the two.[9]

9 Both technical and nontechnical discussions of these methodologies can be found in Kwoka, op. cit. (2015).

21. The relative ease of conducting merger retrospectives using DID-relatively low resource cost, focus on the critical outcome variable, simplicity of conclusions--has led to a proliferation of such studies examining mergers. The best of these are quite informative, and as noted at the outset have recently been compiled in the US and separately in the EU to cast light on the question of the effects of mergers and the effectiveness of merger control. That said, presentations and discussions at the Workshop served the important purpose of highlighting various issues that arise in actually implementing DID. The rest of this note focuses on six such issues that emerged from presentations and discussions in the Workshop.[10] We begin with those arising primarily in the evaluation of mergersBperhaps the most familiar setting of policiesBand then discuss others more distinctive to abuse of dominance cases and, finally, other policy matters.

3.1 Identification of Control Group as a Basis for Policy Analysis

22. Common to most evaluations, and central to the widely used DID methodology, is the need for an appropriate control group for purposes of comparison. Identifying a set of otherwise comparable firms that is, with similar cost and demand experience but unaffected by the merger--is often not straightforward. Firms or products that have the most similar experience may also be those most likely affected by the merger, while those that are more independent may lack comparability. Thus, for example, while clearly subject to the same demand and cost forces, rivals to the merging firms are inappropriate choices since their performance is affected by the merger. The merging firms in other product or geographic markets where they do not overlap might seem to represent a possible control group, but cost changes from the merger might be correlated across markets even if prices are not. Typically, therefore, the control group consists of other firms in analogous markets, requiring careful attention to the degree to which they capture the relevant cost and demand experience.

23. This crucial issue of an appropriate control group was encountered to varying degrees and discussed in all of the case studies presented in the workshop. The reason is that, whether merger or not, each sought to evaluate the policy in question by comparing the performance of the relevant firm with a control group of observations. Perhaps the most straightforward of these case studies was in fact the retrospective of supermarket mergers conducted by US FTC staff. Its control group consisted of geographic markets that did not experience "meaningful change in market structure" during the relevant period, but even this seemingly conventional approach involved an important nuance. The authors identified two different control groups, one comprised of 75 markets without mergers, entry, or exit during the study period, the second a larger group of 117 markets where any such change involved no more than two percent of stores in the market. The latter, more expansive criterion was necessary since the initial control group included no larger markets of the sort that appeared among the treatment group, that is, the markets experiencing mergers.

24. Considerably more challenging are efforts to identify suitable comparison observations for evaluating non-merger cases. The reason is that there is unlikely to be a group of otherwise comparable firms that did not engage in some specific anticompetitive practice, or engaged in the practice but were not subject to the remedy. This difficulty requires more creative ways to achieve some degree of control and comparability. For example, the evaluation of the policy response to strategic withholding by E.ON relied on a creative alternative "control group" against which to measure the effect of the policy on strategic withholding. It compared peak vs. off-peak electricity prices, before and after the measures were taken, on the theory that off-peak prices should not be influenced by strategic withholding (or its absence). To that extent, those prices would seem to represent the "control" group, allowing for the use of something like difference-in-difference. There was discussion at the Workshop about alternative methods for identifying the effects of strategic withholding.

10 Other issues involving the formal econometrics in these evaluations were raised in the Workshop but not summarized here.

25. More challenging yet was the effort to evaluate the effects of the various UK Competition Commission actions against BAA. Identifying a "control" for a monopoly or near-monopoly and associated remedy would seem a logical near-impossibility. Joint ownership of the three Southeast UK airports confronted this issue, as did common ownership of the airports in Glasgow and Edinburgh in Scotland, and the Aberdeen airport itself, given its location. There were, in short, no other airports comparably situated that could serve as counterfactuals for the purpose of evaluating the effects of mandated divestiture of Stansted and Gatwich in the southeast, and either Glasgow or Edinburgh. Faced with this practical problem, the evaluation presented in the Workshop adopted the alternative strategy of compiling data on the various possible effects over time, and observing whether their timing plausibly coincided with divestiture. This approach does not in fact capture the true counterfactual for purposes of DID. Rather, it tests for simple differences (changes, really) and then infers causality based on timing. Under these difficult circumstances, with appropriate caveats, it nonetheless may represent a useful second best approach.

3.2 Selection vs. Exogeneity of the Competitive Event

26. The conventional difference-in-difference methodology implicitly assumes that the event-a merger, say-is exogenously determined. In practice, this means that the event should be a random occurrence, or perhaps more plausibly, does not arise from some characteristic that might be associated with the outcome. In actual practice, of course, a merger is not the pairing of two random firms in an industry, but rather the deliberate choice of firms that perceive some distinctive cost, revenue, or other strategic advantage to their merger. This raises concern that the effect measured in a retrospective in fact is also the result of the same underlying factor that prompted the merger in the first place. If so, this can give rise to incorrect inferences about the effects of mergers.

27. This Workshop had the benefit of one case study that directly and constructively addressed this concern over selection, namely, the FTC staff study of supermarket mergers. As already noted, that study offered two alternative control groupsBone that maximized independence from the treatment category, the other a slightly expanded version that included observations of larger markets that were represented in the treatment group. Altering the control group to better ensure comparability is a simple expedient to address, at least partially, concerns about comparability that can be observed through inspection of group characteristics. But this study is notable for its use of two more sophisticated approaches to this concernBpropensity score matching and synthetic control group. Propensity score matching addresses the selection issue through the use of measures of the likelihood that each firm engages in a merger, whereas synthetic control group analysis develops an entirely artificial or hypothetical control group based on the relevant characteristics. Both of these methods are used in the FTC staff study, and the results compared to those when no such corrections are employed, making this study of distinctive importance.[11]

28. Given that the other two case study evaluations in this Workshop lacked true control groups at all, concern over possible selection issues was not relevant. If the policies toward E.ON or BAA were to be subjected to more explicit DID, however, then the same issues about endogeneity of their conduct and selection of control group observations would need to be addressed.

11 As was noted at the Workshop, it is surprising that this study found little difference in the results with any of these corrections vs. the absence of any correction at all.

3.3 Uniqueness of Alleged Abuses of Dominance

29. As discussed in the Workshop and suggested by these last comments, the studies evaluating policies toward abuse of dominance face additional methodological hurdles. One of the most difficult is the fact that each type of alleged abuse, each individual experience with such abuse, and each policy directed at the abuse experience may be distinctive, perhaps to the point of being unique. Furthermore, distinctiveness is an obstacle to identifying a control group of observations and therefore to the application of DID to alleged anticompetitive practices.

30. Distinctiveness arises, perhaps most obviously, if the action and associated policy involve a truly dominant firm, since then there is almost by definition no other comparable dominant firm, certainly not in the same market. Distinctiveness may also arise with respect to the behavior in question. Such fundamentally distinct methods of abuse as strategic pricing or incompatibility or withholding cannot be aggregated or compared; rather, each must be analyzed separately. Moreover, even within any single category of abuseBsuch as strategic incompatability-individual experiences may well be nearly unique, since each may depend crucially on facts specific to that experience. Similar distinctiveness is likely to characterize policy responses, including remedies, to each such event or action.

31. The different issues raised by distinctiveness are apparent in the policy evaluations of E.ON and of BAA that were presented and discussed in this Workshop. In both of these cases, the setting of the entity and the nature of the allegation against it defied any effort to identify a comparable control group. In the case of E.ON, the relevant comparison would have had to be another large energy supplier that also engaged in strategic withholding but against which there was no remedy of the sort employed by the EU. The study in fact describes various possible benchmark firms that were considered, but rejected. In the case of BAA that entity owned all three southeast UK airports, as well as the airports in Glasgow and Edinburgh. As essentially a monopolist, there were no other airports comparably situated that could serve as counterfactuals for the purpose of evaluating the effects of mandated divestitures.

32. These particular difficulties do not arise as frequently in evaluations of horizontal mergers. Their fundamental similarity makes it possible to identify control group cases and to aggregate experiences in a meaningful way.

3.4 Confounded Causality in the Face of Multiple Overlapping Actions

33. Ideal circumstances for DID-type comparisons involve discrete events and policies, isolated in time and space for purposes of establishing causation. By contrast, overlapping or simultaneous events make such comparisons difficult. For example, a merger by two firms (or a policy taken in response to their merger) that are involved in other mergers around the same time, in an industry that is otherwise undergoing structural changes, and where possible control group firms are also merging poses considerable analytical difficulties. While many retrospectives on mergers and merger policies are able to identify sufficiently isolated cases, this is often much more difficult in the case of allegations of dominant firm abuse. In many-indeed, perhaps mostBcases of anticompetitive conduct, the dominant firm does not employ a single strategy, but rather engages in multiple actions, to achieve its anticompetitive purposes.

34. The resulting difficulties were illustrated by the case studies presented and discussed in this Workshop. As noted at the outset, E.ON faced two allegationsBstrategic withholding of capacity, and preferential use of its own source of balancing power. Both were alleged to occur over the same time period, and while each was believed remedied by a somewhat different policy action, their overlapping time frames and effects made it difficult to disentangle the two. In particular, divestiture simultaneously reduced E.ON's incentive to engage in strategic withholding, and also diminished its possible role in the balancing market. It was therefore difficult to determine which of these mechanisms was ultimately more responsible for

constraining prices, as was found. The potential for confounded causality might affect the choice of competition policy for evaluation or the method to evaluate it.

35. Similar issues, though of lesser importance, presented themselves in the other evaluations presented in the Workshop. The UK policy actions against BAA were largely structural in nature, but overlapping in time. Tracing the causal effect of any single remedial action under these conditions proved quite difficult. In the case of supermarkets, the initial control group was free of other mergers, entry, and exit, but the study was forced to compromise on this in order to improve comparability, as previously discussed. While this strategy seems appropriate, it might haveBbut appeared not to haveBresulted in confounded causality.

3.5 Issues of Timing

36. As discussed by the presenters and commenters in the Workshop, application of DID or any other method for evaluating competition policy actions requires attention to issues of timing. Three different dimensions of timing were identified. The first concerns the point in time after a merger or other event or policy when its effects should be sufficiently well established to permit evaluation. Many retrospectives examine firm or market performance two or three years after the event or policy, a period of time plausibly long enough for firms to adjust, but not so long as to overlap with other events and policies that would complicate the analysis. The exact interval likely differs among firms and policies. A second timing issue is that adjustment to a policy may be gradual rather than abrupt. This argues for an empirical model that is more sophisticated than simply "before and after," in order to correctly measure the actual effects. The third issue--the appropriate timing of any evaluation-will be discussed in the next section.

37. These various issues of timing arose in the three case studies in the Workshop, although in somewhat different forms. The FTC retrospective on supermarket mergers, for example, collected data for 2004-09, and focused on mergers that occurred in either 2007 or 2008 for which there were ten quarters of data. There was little explanation for these choices of intervals-either the overall period or the (apparently) relatively brief postmerger interval for some merger events. The evaluation of policy taken with respect to E.ON adopted an interesting two-interval approach. A shorter post-policy time frameBweeks--was explicitly justified by the argument that energy spot prices were determined quickly in an auction market, while a longer multi-year time frame was also examined in the belief that it might better capture the comparative statics after full adjustment. Quantitative analysis of the effects of UK policies toward BAA were generally conducted for the period 2005-15, bracketing the diverse points in time for the airport divestitures and other remedial measures required of BAA. Complicating the analysis was the fact that these different policies might take varying lengths of time to manifest themselvesBstructural changes longer than certain regulatory and behavioral rules, for exampleBand further, that some very short term effects might be perverse due to the disruption associated with changes in ownership, operations, and physical infrastructure.

38. There was agreement at the Workshop that further explicit attention to the criteria for the choice of timing would strengthen the analysis and conclusions of these evaluations of competition policy initiatives.

3.6 Effects on Product or Service Quality

39. Workshop participants noted that the disproportionate attention given to the effects of policy on price, with less--or in some cases no--attention to product or service quality. While quality might not be a dimension of equal importance in some cases, research seemed to emphasize that which was readily quantifiable even when less quantifiable quality issues were arguably important. This characteristic of competition evaluation-and criticisms of it-are commonplace in the literature, and were reflected in the policy evaluations presented in the Workshop.

40. In this connection, the EU evaluation of actions taken toward BAA was notable for its considerable attention to the quality dimensions of air travel, airports, and air servicing at the relevant airports. It specifically addressed such factors as the number of seats and seats filled; route development and churn; airline switching; infrastructure improvements affecting ground transport, immigration control, and baggage handling; and measured passenger satisfaction. This emphasis in the study was rightly regarded as exemplary, but also illustrated the inherent difficulties of this pursuit. The multidimensional nature of "quality," the lack of agreed upon metrics and data, and the concern as to whether improvements were worthwhile limited the ability of this study to arrive at a definitive conclusion, but its effort and approach served as a very useful complement to the emphasis on price that guides most evaluations.

41. The other two evaluationsBof US policy toward supermarket mergers and the EU policy toward E.ON's behavior-focused on resulting prices, although these seemed to be cases where that emphasis was appropriate.

5. Resources and Procedures

42. As noted at the outset, there was a consensus view that ex post evaluations are valuable policy tools, but it was also clear that there was concern regarding the resources requirements for such evaluations. Several considerations were raised. Perhaps the most immediate is the direct cost of performing an evaluation, whether done internally within a competition agency or externally by a researcher or research center. The first scenario is lower cost only if there is slack within the agencyBthat is, if such an undertaking coincided with a period where enforcement activity was modest. Since few agencies have or can count on such slack, committing the necessary internal resources would affect on-going enforcement activities. While externally commissioned work avoids that problem, the costs are explicit and invite scrutiny regardless of the importance of the evaluation.

43. The other cost that was discussed several times involved acquisition of necessary data. While the precise cost depends on the specific data, of course, the cost of data used in one of the evaluations presented at this Workshop was approximately USD 100,000. Another evaluation project cost EUR 200,000 for both data and outside consulting services. Some delegates noted that such amounts would be difficult to secure in their agencies.

44. A further procedural matter that also attracted considerable comment involved the timeline for any evaluation and resulting report, and also where in that timeline would be the most useful point to interject comments and critiques. With respect to the timeline, some participants noted an emphasis on evaluating policy actions fairly soon after they were implemented. While this emphasis served a perceived "need to know," it was also noted that it might be at odds with the somewhat longer interval before full ex post effects emerged as well as simply the time requirements of reliable research. Finally, there was considerable agreement that comments, suggestions, and critiques of these evaluations, such as offered in this Workshop, would be more valuable earlier in the process. The obvious reason for this is that earlier comments would benefit and improve the research itself, rather than serving largely as an ex post evaluation of the ex post policy evaluation.

6. Conclusions

45. This Workshop illustrated and highlighted numerous aspects of ex post policy evaluations. It certainly demonstrated the widely held conviction in their value, even as it raised questions and concerns about the cost of undertaking such evaluations. It illustrated some ambitious objectives, including the application of evaluation methodology to abuse of dominance cases and to divestiture remedies as well as to the more conventional setting of mergers. It demonstrated applications to a range of quality outcomes and nonprice effects generally, filling in gaps in the literature. From a methodological perspective, the evaluations included cases that lacked clear counterfactuals and required alternative methods for isolating effects due to the policy actions. And they illustrated state of the art empirical techniques to control for selection issues that have received too little attention in the past, thereby ensuring that selection is not responsible for the reported results.

46. This list of contributions, both substantive and methodological, is quite substantial. It confirms the value of continuing to address new policy questions and to strengthen the methodologies being used. These advances bode well for the common interest in improving competition policy actions.