

(MIS)RELIANCE ON SOCIAL SCIENCE EVIDENCE IN INTELLECTUAL PROPERTY LITIGATION: A CASE STUDY*

*E. Richard Gold and Robert Carbone***

ABSTRACT

Although courts increasingly rely on social science evidence in litigation, this practice has undergone surprisingly little analysis outside Charter litigation. In particular, although social science evidence—including surveys, economic studies, and estimates of market losses—is commonplace in intellectual property litigation, it has attracted only limited substantive or critical analysis. Because this evidence is fundamental not only to questions of liability, particularly in trade-mark litigation, but also to damages across all forms of intellectual property, it is time to fill this gap. This article begins the discussion by pointing to the urgent need for judges and lawyers to be discerning in their assessment of the reliability of this evidence. It then illustrates the difficulties of relying on social science evidence in intellectual property litigation by examining the admissibility of a study conducted by the Patented Medicine Prices Review Board (PMPRB) that has potential to be used in a patent case. On the basis of this examination, the authors propose guidelines to assist courts in making determinations concerning the admissibility of social science studies.

RÉSUMÉ

Alors que les tribunaux comptent de plus en plus sur la preuve d'ordre social dans les litiges, il est surprenant de constater que cette pratique a fait l'objet de très peu d'analyse en dehors des litiges concernant la *Charte*. En effet, bien que la preuve d'ordre social, notamment les enquêtes, les études économiques et les estimations des pertes marchandes, soit couramment utilisée dans les litiges en matière de propriété intellectuelle (PI), celle-ci a attiré un nombre limité d'analyses de fond ou de critiques. Comme cette preuve est fondamentale non seulement au regard des questions de responsabilité, particulièrement dans les litiges en matière de marque de commerce,

* Submission to the Editor, June 11, 2012.

** © 2012 E. Richard Gold, S.J.D. (Michigan), James McGill Professor, Faculty of Law, McGill University; and Robert Carbone, Ph.D. (Carnegie Mellon), President, Futurion Inc. The authors would like to thank the research assistance of James Duffy and the editorial assistance of both James Duffy and Julia Carbone. The authors also wish to acknowledge the helpful comments of two anonymous reviewers who reviewed an early draft of this article. This research was supported, in part, by VALGEN (Value Addition through Genomics and GE3LS), a project sponsored by the Government of Canada through Genome Canada, Genome Prairie, and Genome Quebec.

mais aussi en ce qui a trait aux dommages-intérêts pour toutes les formes de PI, le temps est venu de combler cette lacune. L'article débute en notant le besoin urgent pour les juges et les avocats de faire preuve de discernement dans leur évaluation de la fiabilité de cette preuve. Les difficultés du recours à la preuve d'ordre social dans les affaires de PI y est ensuite illustré en examinant la recevabilité d'une étude réalisée par le Conseil d'examen du prix des médicaments brevetés susceptible d'être utilisée dans une affaire en matière de brevet. Sur la base de cet examen, les auteurs proposent des lignes directrices pour aider les tribunaux à prendre des décisions relativement à la recevabilité des études en sciences sociales.

CONTENTS

1.0	Introduction	180
2.0	A Case Analysis: The Patented Medicines Price Review Board Study	182
2.1	Background	183
2.2	Limitations of the PMPRB Study	184
2.3	Methodology	185
2.4	Arbitrary Application of Methodology	189
2.5	Data Disclosure	191
2.6	Peer Review	192
3.0	A Practice for the Admission of Scientific Studies	193
4.0	Conclusions	195

1.0 INTRODUCTION

Although courts increasingly rely on social science evidence in litigation, this practice has undergone surprisingly little analysis outside Charter litigation.¹ In particular, although social science evidence—including surveys, economic studies, and estimates of market losses—is commonplace in intellectual property litigation, it has attracted only limited substantive or critical analysis.² Because this evidence is fundamental not only to questions of liability, particularly in trade-mark litigation,³ but also to damages across all forms of intellectual property, it is time to fill this gap. The general principles for admission of expert evidence in Canadian criminal and civil cases are set out in *R. v. Mohan*⁴ as elaborated in *R. v. J.-L.J.*⁵ and *R. v. Trochym*.⁶ These rules apply regardless of whether the trier-of-fact

¹ See e.g. Christopher P. Manfredi & James B. Kelly, “Misrepresenting the Supreme Court’s Record? A Comment on Sujit Choudhry and Claire E. Hunter, ‘Measuring Judicial Activism on the Supreme Court of Canada’” (2004) 49 McGill L.J. 741, online: McGill Law Journal <http://lawjournal.mcgill.ca/documents/Manfredi_and_Kelly.pdf>.

² See *Masterpiece Inc. v. Alavida Lifestyles Inc.*, 2011 SCC 27, [2011] 2 S.C.R. 387 [*Masterpiece v. Alavida*], in which expert evidence was rejected because it was found unnecessary and unreliable.

³ *Ibid.*

⁴ [1994] 2 S.C.R. 9 [*Mohan*].

⁵ 2000 SCC 51, [2000] 2 S.C.R. 600 [*J.-L.J.*].

⁶ 2007 SCC 6, [2007] 1 S.C.R. 239 [*Trochym*].

is a jury or judge. They thus set out the test for the inclusion of expert evidence in intellectual property litigation.

According to both Sopinka J. in *Mohan* and Binnie J. in *R. v. J.-L.J.*, the judge is entrusted with keeping “junk” science—no matter whether that *science* originates in the natural, life, or social sciences—out of the courtroom.⁷ As Justice Stephen Breyer of the Supreme Court of the United States aptly explains,⁸ it is the judge who is charged with excluding studies that, although purporting to be scientific, are based on faulty methodology or a lack of transparency or otherwise fail to meet the basic requirements of reliability:

I am reminded of a remark made by the physicist Wolfgang Pauli. After a colleague asked whether a certain scientific paper was wrong, Pauli replied, “Oh, no. Certainly not. That paper is not good enough to be wrong.” That is our objective. It is to avoid legal decisions that reflect that paper’s so-called science. It is to avoid the kind of serious scientific mistake that once led one court, for example, to hold that dropping a can of orange juice caused cancer.⁹

Justice Deschamps in *R. v. Trochym* summarized the Court’s conclusions regarding the admissibility of scientific evidence as follows:

In *J.-L.J.*, Binnie J. explained that Canadian courts require a “reliable foundation” for novel science to be admissible as evidence at trial. Drawing on the American case of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹⁰ he observed that reliability can be evaluated on the basis of four factors (*J.-L.J.*, at para. 33):

- (1) whether the ... technique can be and has been tested[;] ...
- (2) whether the ... technique has been subjected to peer review and publication[;] ...
- (3) the known or potential rate of error ... ; and,
- (4) whether the theory or technique used has been generally accepted.¹¹

Despite the clear gatekeeping role that the trilogy of cases—*Mohan*, *J.-L.J.*, and *Trochym*—imposes on the judge, neither case law nor academic articles provide any substantive guidance to judges on the process through which they ought to

⁷ See e.g. *J.-L.J.*, *supra* note 5 at para. 25:

However, the dramatic growth in the frequency with which they have been called upon in recent years has led to ongoing debate about suitable controls on their participation, precautions to exclude “junk science,” and the need to preserve and protect the role of the trier of fact—the judge or the jury.

⁸ Building on the U.S. Supreme Court decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993) [*Daubert*], which Binnie J. stated sets out “a number of factors that could be helpful in evaluating the soundness of novel science” in Canada, including whether the study constitutes “good science”: *R. v. J.-L.J.*, *supra* note 5 at para. 33.

⁹ Stephen Bryer, “Interdependence of Science and Law” (1998), 89 *Judicature* 24 at 25 [Bryer].

¹⁰ *Supra* note 8.

¹¹ *Trochym*, *supra* note 6 at para. 36.

evaluate whether a purported social science study is sufficiently reliable to be admitted. In the field of intellectual property, the clearest statement on this subject is found in *Masterpiece v. Alavida*. In that case, Rothstein J. refused to admit a social science survey because of its lack of reliability: “For a survey to be valid, it seems elementary that there must be some consumers who could have an imperfect recollection of the first mark. Simulating an ‘imperfect recollection’ through a series of lead-up questions to consumers will rarely be seen as reliable and valid.”¹²

What makes a social science study sufficiently reliable to constitute “good science”? Is it appropriate for a judge to admit a social science study where the authors of the study are not present to undergo cross-examination? How transparent must the study be in explaining its methodology, data, assumptions, and limitations before a judge should admit it? Is it appropriate for a judge to take judicial notice of a study that is publicly available? Although these are critical questions in general, they have significant relevance in intellectual property disputes where social science studies play a critical role in assessments of damages.

2.0 A CASE ANALYSIS: THE PATENTED MEDICINES PRICE REVIEW BOARD STUDY

In order to illustrate the difficulties of relying on social science evidence in intellectual property litigation, this article examines the potential use in Canadian patent litigation of a study that the Patented Medicine Prices Review Board (PMPRB) published in 2011 and revised in 2012.¹³ Among other things, the study may be relevant in determining the size of the market for a pharmaceutical product under either s. 55 of the *Patent Act*¹⁴ or s. 8 of the *PM(NOC) Regulations*.¹⁵ Because the source of the study, the PMPRB, is an independent governmental body whose work, it would normally be assumed, meets accepted practice, the failure of the study to meet this standard points to the serious danger of reliance on social scientific or scientific study without evaluating whether it is “good enough to be wrong.”

The PMPRB undertook the study¹⁶ as part of the National Prescription Drug Utilization Information System (NPDUIS). The PMPRB study examines the effect of

¹² *Supra* note 2 at para. 96.

¹³ The discussion is no longer hypothetical. This article was prepared before the release of reasons in *Apotex Inc. v. Sanofi-Aventis*, 2012 FC 553 on May 23, 2012 [*Apotex v. Sanofi-Aventis*]. In that case, the judge discussed the PMPRB study at length and relied on it to prefer one expert’s report over the other: *ibid.* at para. 94. We note that one of the co-authors, Dr. Carbone, was an expert witness in that case.

¹⁴ R.S.C., 1985, c. P-4.

¹⁵ *Patented Medicines (Notice of Compliance) Regulations*, S.O.R./1993-133.

¹⁶ *The Impact of Generic Entry on the Utilization of the Ingredient* (revised May 2012) (Ottawa: Patented Medicine Prices Review Board, 2012) [PMPRB study], online: Patented Medicine Prices Review Board <<http://www.pmprb-cepmb.gc.ca/CMFiles/Publications/Analytical%20Studies/NPDUIS-GenericEntryImpact-REDO-e-may7.pdf>>. The original version of the report from September 2011 is no longer available online. We have a copy of this report on file. Although authorship of the study is not provided, the study states that the research was conducted by the PMPRB.

the entry of generic versions of seven patented pharmaceutical compounds on the use of those compounds exclusively within public drug plans in six Canadian provinces. It concludes, based on the data from these public plans, that there is no decrease in the use of the drug compound after generic entry into the market. Not only is this result in direct contradiction of a substantial body of existing evidence,¹⁷ but its methodology has serious flaws. Thus, if this report were to be considered in determining damages, it could lead to an unjust result.

Given the lack of clear practice or analytical framework for determining whether the judge ought to admit into evidence a particular social science study, it is possible that the PMPRB study could be improperly relied on, even if indirectly, in intellectual property litigation.¹⁸ In an infringement action, this would mean that the losses suffered by the patent holder could be calculated on the basis of the actual total volume of the compound consumed during the period following the wrongful entry of an infringing generic product. In an action under s. 8 of the *PM(NOC) Regulations*, this would mean that the volume of product that the generic company would have sold could be calculated on the basis of the patent holder's actual sales during the prohibition period. In the first situation, the court would be led to significantly underestimate damages whereas, in the latter, the court would be led to significantly overestimate damages.

In order to help develop practice aimed at keeping out studies that are “not good enough to be wrong,” we provide a critique of the PMPRB study and explain why its use would be inappropriate in patent litigation. After providing background information on the PMPRB study, we focus on five elements: the limitations of the PMPRB study; its methodology; the arbitrary application of the methodology; the failure to disclose the data relied on; and the absence of external peer review. Based on this critique, this article concludes by proposing guidelines for the assessment of the admissibility of social science studies into evidence.

2.1 Background

The PMPRB study was originally released in September 2011.¹⁹ Because of methodological concerns raised by the authors of the present article,²⁰ the PMPRB revised

¹⁷ Peter J. Huckfeldt & Christopher R. Knittel, “Pharmaceutical Use Following Generic Entry: Paying Less and Buying Less” (2011), NBER Working Paper 17046, online: National Bureau of Economic Research <www.nber.org/papers/w17046>; Richard E. Caves, Michael D. Whinston, & Mark A. Hurwitz, “Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry” (1991) *Brookings Papers: Microeconomics* 1991 at 1; Darius Lakdawalla, Tomas Philipson, & Y. Richard Wang, “Intellectual Property and Marketing” (2006) NBER Working Paper 12577, online: National Bureau of Economic Research <www.nber.org/papers/w12577>.

¹⁸ As discussed in Bryer, *supra* note 9, this has, in fact, occurred.

¹⁹ PMPRB, *The Impact of Generic Entry on the Utilization of the Ingredient* (Ottawa: Patent Medicine Prices Review Board, 2011) [“the 2011 study”]. This study is on file with the authors.

²⁰ These included the failure to correctly describe the methodology. In the September 2011 version of the report, the study's authors stated that their “[p]rojections were made using Microsoft Excel's

the 2011 study in May 2012 to make its methodology more transparent. Moreover, it changed some of the historic data used as well as trend projections, but maintained its overall conclusion that generic entry did not affect the volume of drug compounds consumed in the public system.

The PMPRB study's core analysis consists of developing "but for" forecasts to determine the likely volume of the selected drug compounds that would have been used had there never been generic entry. (The PMPRB study also included long-term analyses that did not involve any forecasts.) Such but-for trend projections are established on historical data observed before an intervening event that actually occurred. For the most part, the PMPRB study authors relied on 12 months of data but, on occasion, used 15 and 24 months of historical data prior to generic entry. The analysis focused on seven top-selling drugs—omeprazole, risperidone, simvastatin, pravastatin, citalopram, paroxetine, and gabapentin. Each drug lost patent protection in Canada in the last decade. Two of the drugs are specialty products (an antipsychotic for the treatment of schizophrenia and a drug to treat epilepsy). The remaining five are commonly prescribed by general practitioners. Use was determined based on the number of claims made for the compound to a provincial drug plan. Other information is presented in the report, but is not relevant to the current discussion.

2.2 Limitations of the PMPRB Study

The PMPRB study itself warned against using its data or analysis to draw any conclusions with respect to the Canadian pharmaceutical market as a whole. The study's authors made clear that they drew only on data from *public* drug plans in *six* provinces for the short-term analyses (eight provinces were included in the long-term analyses but these were not used as the basis of any but-for forecast²¹) and thus were not representative of the entire market:

Furthermore, public drug plan data represents only one component of the overall pharmaceutical market. Therefore, these results should not be extrapolated to the overall Canadian marketplace, as the impact may be different in markets reimbursed by private insurers or drugs paid for out-of-pocket by patients.²²

Limitations are easy to overlook, because they are seldom included in executive summaries and are often set out only in footnotes or otherwise isolated from the main body of the text.²³ In this case, the PMPRB study's limitations make it clear

base forecasting function, which uses a linear regression model to calculate future values," *ibid.* at 3, whereas it was apparent that they had more often used nonlinear regression models. Further, while the 2011 study's authors stated that "the short-term assessments were reported based on 12-month periods prior to and post generic entry," *ibid.* at 3, they actually used various ranges. There were other, less significant, failures to fully disclose the methodology used.

²¹ The PMPRB study's authors relied on the long-term analyses for visual examination of yearly use patterns both pre- and post-generic entry.

²² PMPRB study, *supra* note 16 at 4.

²³ Presumably, this is what happened in *Apotex v. Sanofi-Aventis*, *supra* note 13, where this limitation seems not to have been drawn to the judge's attention. The judge incorrectly stated, *ibid.* at 93, that

that it would be inappropriate for it to be admitted into evidence and used to calculate damages across the general Canadian pharmaceutical marketplace.

2.3 Methodology

The PMPRB study sets out a general description of the methodology used in conducting the analysis. Data was obtained from the NPDUIS database for Alberta, Saskatchewan, Manitoba, New Brunswick, Nova Scotia, and Prince Edward Island for short-term analysis. (The long-term analyses relied on these six provinces as well as data obtained from Ontario and British Columbia.²⁴) Given that the utilization conclusions were based on only the short-term analysis, the PMPRB study did not consider the use of drug compounds in Canada's three most populous provinces: Ontario, Quebec, and British Columbia.

The PMPRB study's counterfactual projections of the level of use of drug compounds in the absence of generic entry relied on a variety of different trend models "including linear, logarithmic, power, exponential and polynomial based forecasts."²⁵ While the PMPRB study stated that it generally employed a linear trend model, it specified that other models were selected in the following cases:

1. where reductions in market shares exhibited a convex curvature prior to generic entry;
2. where markets exhibited seasonality; and
3. where there were limited and fluctuating observations prior to generic entry.²⁶

In all of these cases, the PMPRB study's authors employed a logarithmic trend model.

The selection of a trend model has a significant impact on any projection and thus the use of the wrong model can seriously bias results.²⁷ A linear model provides a middle path between assuming that trends become weaker over time (level out)—the assumption behind a logarithmic model—or strengthen (explode) as assumed by an exponential model. A linear trend model would predict that the growth in the use of a drug compound would be constant over time. A logarithmic trend model would predict that the entire market for a compound would become saturated—that

"[t]he purpose of the study carried out by the PMPRB was to determine whether, upon genericization, a drug continues to be utilized to the same extent. This is the exact question that [the experts] addressed in their reports." As the PMPRB study's authors made clear, this was not the purpose of the study, which was restricted to the use of compounds within *public* drug plans, and the study's authors specifically warned that the study should not be used to extrapolate any conclusions to the entire Canadian pharmaceutical market.

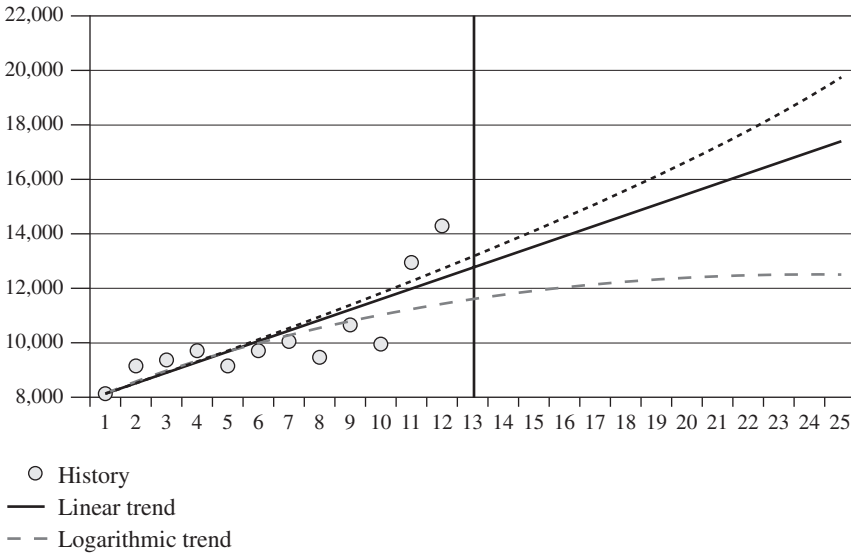
²⁴ PMPRB study, *supra* note 16 at 2.

²⁵ *Ibid.*

²⁶ *Ibid.*

²⁷ Paolo Brandimarte, *Quantitative Methods: An Introduction for Business Management* (Hoboken, NJ: John Wiley & Sons, 2011) at 555: "However, bias can also be the effect of a wrong demand model" [Brandimarte].

Figure 1 Comparing Linear, Exponential, and Logarithmic Trend Lines Using 12 Data Points



is, it assumes that there is little market growth potential for the compound. An exponential trend would predict that the growth in use is accelerating with ever-increasing use of the compound.²⁸ Figure 1 illustrates these different trend models.

“[L]inear models are the natural starting point”²⁹ for the development of trend projections. If one deviates from the linear model, one needs to justify doing so based on either (1) existing knowledge—theoretical and empirical—in the field³⁰ or (2) the fact that the other model provides a better fit with the data.³¹ The PMPRB study does not attempt to provide either type of justification for the selection of a logarithmic trend model.

First, the selection of a logarithmic trend model is not justified by existing theory or empirical knowledge. Further, it directly contradicts what is known in the field. A significant number of studies have concluded that generic entry has a nega-

²⁸ Mathematically, the logarithmic function is the mirror image of an exponential function. In the case of a positive trend, the curvature of the latter function is convex, exhibiting, first, slow growth and, then, exploding values, whereas the curvature of the former function is concave, displaying, first, rapid increases and, then, levelling-out values.

²⁹ Brandimarte, *supra* note 27 at 479.

³⁰ John H. McDonald, *Handbook of Biological Statistics*, 2d ed. (Baltimore, MD: Sparky House Publishing, 2009) at 149-50; Holger Kantz & Thomas Schreiber, *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press, 1997) at 3.

³¹ Brandimarte, *supra* note 27 at 505: “The R^2 coefficient gives us an evaluation of the overall fit of the model.”

tive effect on the continued use of a drug compound.³² A paper released four months before the 2011 study concluded, in fact, “that use of medications falls in the years immediately before and after generic entry, and that this happens even when prices fall or are stable in real terms.”³³ The authors of that paper supplemented their regression analyses with case studies of various compounds including two examined in the PMPRB study: simvastatin and citalopram. One reason that use of medications falls is that pharmaceutical companies switch their marketing efforts from the drug that became subject to generic entry to reformulated drugs still under patent protection: “[W]e also see that the utilization of the branded molecule begins to fall prior to generic entry, and that this coincides with increased utilization of branded reformulations.”³⁴ There is significant evidence that pharmaceutical companies spend heavily on marketing, with one recent study concluding that the industry spends almost twice as much on marketing as on research and development.³⁵ This spending provides an explanation why, despite the entry of generics, overall use of the drug drops. Marketing of reformulations and other competing branded products changes prescribing behaviour despite the lower prices of the original (off-patent) compound.

One of the reasons the PMPRB study’s authors advocated use of a logarithmic trend model was seasonality. However, the literature does not suggest the use of a logarithmic model to deal with seasonality. Rather, seasonality in trends is generally dealt with through additive or multiplicative decomposition in which the underlying trend and each of a seasonality and an error factor are either added together or multiplied.³⁶ Alternatively, seasonality can be taken into account through some form of exponential smoothing or other methods.³⁷ Notably, the imposition of a logarithmic

³² See articles cited *supra* note 17.

³³ Huckfeldt & Knittel, *supra* note 17 at 19.

³⁴ *Ibid.*

³⁵ Marc-André Gagnon & Joel Lexchin, “The Cost of Pushing Pills: A New Estimate of Pharmaceutical Promotion Expenditures in the United States” (2008) 5 PLoS Med, online: PLOS Medicine <<http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050001>>: “From this new estimate, it appears that pharmaceutical companies spend almost twice as much on promotion as they do on R&D.”

³⁶ *Ibid.* at 539.

³⁷ *Ibid.* at 547. Where the seasonal variation is itself increasing—that is, the variation increases year to year—one can use a logarithmic *transformation* to smooth that variation. Unlike a logarithmic trend model, the appropriate logarithmic transformation involves, at a minimum, taking the logarithm of the dependent variable (the utilization of the compound). Thus, the imposition of a logarithmic trend model is very different from the aforementioned logarithmic transformation. In any case, such a transformation is only appropriate where there is evidence that the seasonal variation is increasing. Bruce L. Bowerman, Richard T. O’Connell, & Anne B. Khoeler, *Forecasting, Time Series, and Regression*, 4th ed. (Belmont, CA: Thomson Brooks/Cole, 2005) at 295-96. Given that the short-term analysis involves 12 months (and at most, 24 months) of data, it is not possible to determine whether the seasonal variation is actually increasing. It is thus inappropriate to use a logarithmic *transformation* in this case.

trend model, as employed in the PMPRB study, is neither a standard nor referenced method for adjusting for seasonality.

If the PMPRB study did not follow theory, it also did not select models based on best fit. In statistics, fit is commonly determined by calculating the coefficient of determination. This coefficient essentially determines how closely the trend line matches the actual historical data. The coefficient ranges from 0 to 1 with the best fit being as close to 1 as possible. In examining the coefficient of determination for each of the various projections used in the PMPRB study, one quickly finds that the PMPRB study does not use the trend line with the highest coefficient. For example, the PMPRB study's authors selected a logarithmic trend for pravastatin in Alberta. Yet, the coefficient of determination is highest when an exponential trend model is used, as depicted in Figure 2, and *lowest* for the PMPRB study's selected logarithmic model. In fact, in no case did the PMPRB study employ an exponential model, even when it resulted in the highest coefficient of determination. Although a coefficient of determination is not the only measure for how closely a curve fits data,³⁸ it is standard practice to present coefficient values and to justify why a curve with a lower coefficient was used. In fact, the PMPRB study itself stated that the coefficient of determination was a criterion for selection of the most appropriate trend.

Given their lack of justification for selecting logarithmic trend models in several key areas of the PMPRB study, the study's authors departed from practice in not uniformly using a linear trend model. Such a trend model would have shown that generic entry had a *negative* effect on the utilization of the compounds being studied, in line with previous studies.³⁹

Another set of methodological problems arises with the PMPRB study's authors' choice to use different lengths of historical data on which to base their projections. Although the PMPRB study stated that, as a general rule, it would use 12 months of data, nevertheless, "in cases for which the 12-month period was insufficient to reflect overall trends in utilization, up to 24 months of data were considered to ensure accuracy."⁴⁰ However, the authors provided no explanation for what would be considered "insufficient." Further, they gave no justification for not uniformly using the period of 24 months in all cases. Second, given the earlier literature that concluded use rates fell "in the years immediately before and after generic entry" because of marketing changes,⁴¹ the PMPRB study authors' selection of between 12 and 24 months of data prior to generic entry risked distorting the results obtained. Earlier scholar-

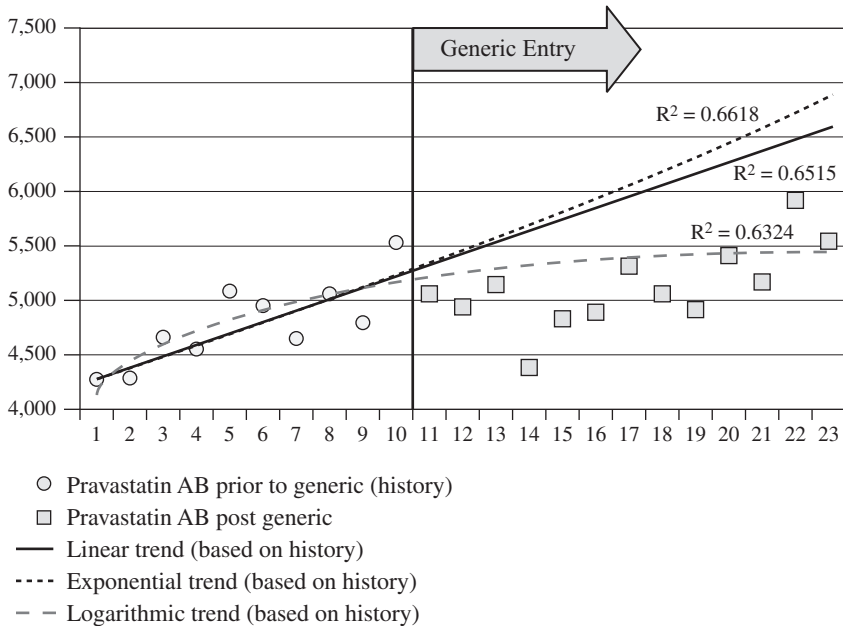
³⁸ Harvey Motulsky & Arthur Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting* (Sand Diego, CA: GraphPad Software, 2003) at 134-36.

³⁹ The authors calculated the trend projects based on reversed-engineered data. See discussion, *infra*, under heading 2.5, "Data Disclosure."

⁴⁰ PMPRB study, *supra* note 16 at 3.

⁴¹ Huckfeldt & Knittel, *supra* note 17 at 19.

Figure 2 A Comparison of the Fitted and Projected Values Obtained with a Linear, Exponential, and Logarithmic Trend Regression for Pravastatin in Alberta and the Respective Coefficients (R^2) Based on Data Reverse-Engineered from the PMPRB Study



Source: Adapted from PMPRB, *The Impact of Generic Entry on the Utilization of the Ingredient* (revised May 2012) (Ottawa: Patent Medicine Prices Review Board, 2012), available online: Patented Medicine Prices Review Board <[http://www.pmprb-cepmb.gc.ca/CMFiles/Publications/Analytical Studies/NPDUIS-GenericEntryImpact-REDO-e-may7.pdf](http://www.pmprb-cepmb.gc.ca/CMFiles/Publications/Analytical%20Studies/NPDUIS-GenericEntryImpact-REDO-e-may7.pdf)>.

ship clearly pointed to the need to draw on significantly longer periods. The decision not to do so is particularly unfortunate given that the data is publicly available.⁴²

Overall, the PMPRB study does not abide by standard practice in conducting trend projections and provides no justification for so doing.

2.4 Arbitrary Application of Methodology

In addition to not abiding by standard practice or justifying its novel methodology, the PMPRB study is inconsistent and arbitrary in its selection of both historical data and trend model. Because most principles of forecasting aim to minimize inconsistency, this is problematic.⁴³ Consider, for example, the discussion in the PMPRB

⁴² Bowerman, O'Connell, & Khoeler, *supra* note 37 at 19.

⁴³ Jon Scott Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Boston, MA: Kluwer Academic, 2001) at 60.

study of the use of paroxetine after generic entry. In the notes accompanying Figure 5.1 of the PMPRB study, reproduced below (see Figure 3), the PMPRB study's authors stated that they selected a linear trend model for all provinces except Manitoba, where they selected a logarithmic trend model.⁴⁴ The stated reasons for doing so were that "[t]he historical 24-month trend in utilization in this jurisdiction exhibits sizable annual variations and diminishing yearly uptake. It was found that the use of a simple linear function would fail to account for the correction in seasonality."⁴⁵ The authors thus selected the logarithmic curve because they claimed that it produced the best fit.

Apart from the lack of consistency among provinces, the selection of a logarithmic trend model, which has the effect of assuming that the use of paroxetine would level out, is perplexing in that the annual variations in Manitoba, as set out in the PMPRB study's Figure 5.5 (see Figure 4), were not substantially different from that of any other province.⁴⁶ Further, there was no explanation for why a logarithmic trend would correct the seasonality issue in Manitoba, but not in Nova Scotia.

While selection of the logarithmic trend model for use of paroxetine in Manitoba was, as suggested by the PMPRB study's authors, based in part on best fit, an analysis of the different trend methods employed does not support this suggestion. This analysis demonstrates that both an exponential and a linear model resulted in a better fit.⁴⁷

For example, a closer examination of Figure 5.1 from the PMPRB study (see Figure 3) reveals a significant *decrease* in use after generic entry. One way to determine whether a change in trend occurs after a specific event in time is to take trend lines of the data before and after the event in question and compare the two.⁴⁸ To illustrate the change in trends, the study's Figure 5.1 (see Figure 3, below) was supplemented by two lines. The first, longer, solid line represents a linear regression of the historical data up to generic entry and a linear projection of that line post-entry. The second, shorter, dotted line represents a linear regression of the historical data from generic entry forward. Normally, the slopes of linear trend lines can be compared using specific mathematical calculations.⁴⁹ However, this option was unavailable because of lack of access to the exact data. Regardless, it can be seen from the modified Figure 3 that the trend is positive (upward) until generic entry and nega-

⁴⁴ PMPRB study, *supra* note 16 at 19.

⁴⁵ *Ibid.*

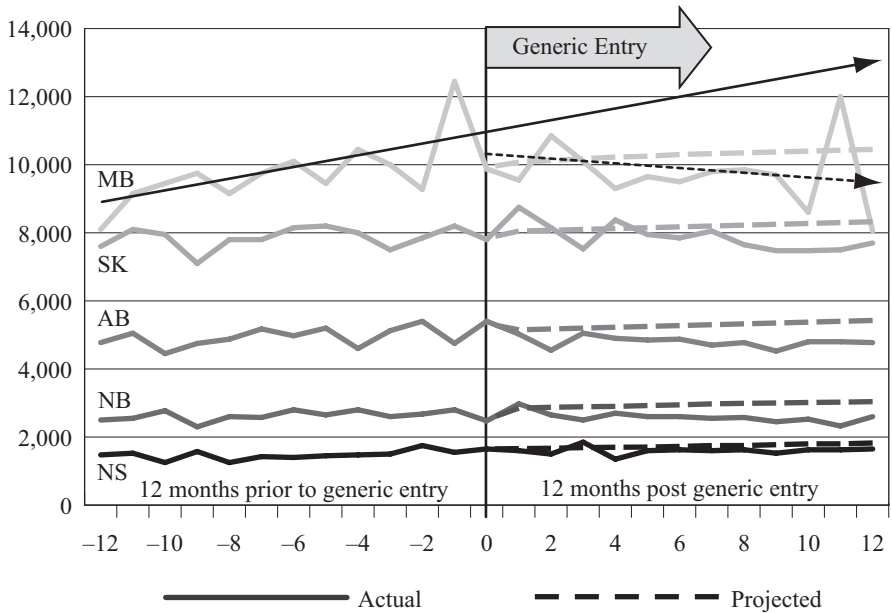
⁴⁶ *Ibid.* at 21.

⁴⁷ See discussion, *supra* note 39. The calculated R^2 values for paroxetine in Manitoba are 0.5126 for the exponential model, 0.4965 for the linear model, and 0.4698 for the selected logarithmic trend model.

⁴⁸ Patrick F. Darken, "Testing for Changes in Trend in Water Quality Data" (unpublished Ph.D. Thesis, Virginia Polytechnic Institute and State University, 1999) at 70.

⁴⁹ Robert R. Sokal & R. James Rohlf, *Biometry*, 4th ed. (New York: W.H. Freeman and Company, 2012) at 513-15.

Figure 3 PMPRB Figure 5.1: Paroxetine—Number of Claims by Jurisdiction*



* Projections are based on a linear function for all jurisdictions except Manitoba. The historical 24-month trend in utilization in this jurisdiction exhibits sizable annual variations and diminishing yearly uptake. It is found that the use of a simple linear function would fail to account for the correction in seasonality. Both the logarithmic and the power functions yielded similar projections, with higher values for the coefficient of determination than the linear function. The logarithmic projection is reported for Manitoba.

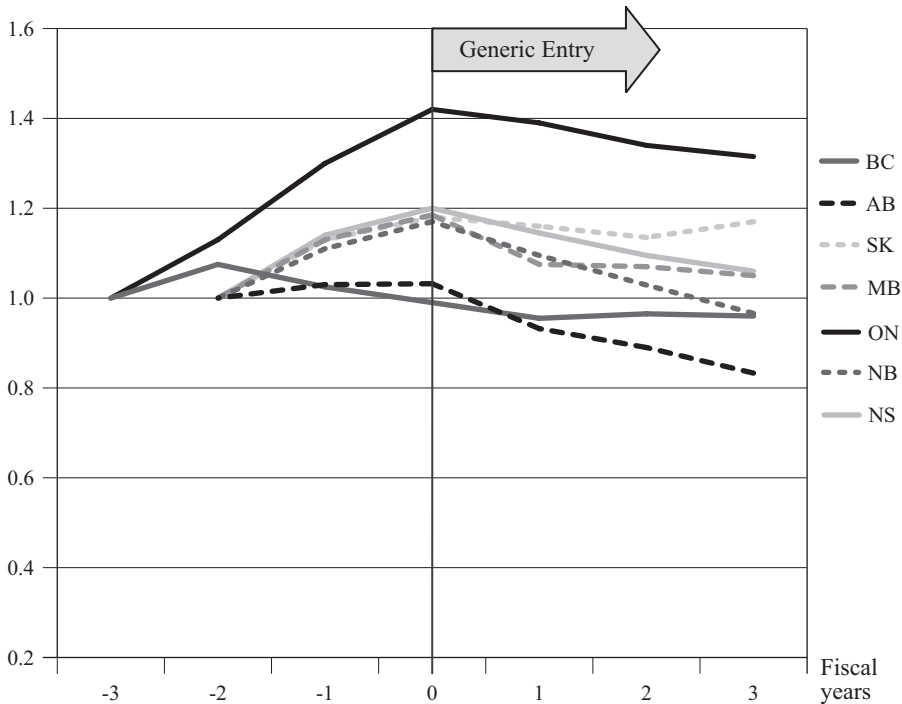
Source: Adapted from PMPRB, *The Impact of Generic Entry on the Utilization of the Ingredient* (revised May 2012) (Ottawa: Patent Medicine Prices Review Board, 2012), available online: Patented Medicine Prices Review Board <<http://www.pmprb-cepmb.gc.ca/CMFiles/Publications/AnalyticalStudies/NPDUIS-GenericEntryImpact-REDO-e-may7.pdf>>.

tive (downward) afterward. This contradicts the PMPRB study’s finding that “the entry of the generic version had little impact on the utilization” of paroxetine.⁵⁰ The PMPRB study’s finding is a result not of the underlying data but of the imposition of a logarithmic model in contradiction to theory, evidence, and practice.

2.5 Data Disclosure

The data on which the PMPRB study’s analysis was based were neither disclosed nor, following requests, made available for review. Although the NPDUIS database is publicly available, it is not possible to reconstruct the data used in the PMPRB

⁵⁰ PMPRB study, *supra* note 16 at 19.

Figure 4 PMPRB Figure 5.5: Paroxetine—Ingredient Utilization Index

Source: Adapted from PMPRB, *The Impact of Generic Entry on the Utilization of the Ingredient* (revised May 2012) (Ottawa: Patent Medicine Prices Review Board, 2012), available online: Patented Medicine Prices Review Board <<http://www.pmprb-cepmb.gc.ca/CMFiles/Publications/Analytical%20Studies/NPDUIS-GenericEntryImpact-REDO-e-may7.pdf>>.

study without knowing the parameters used by its authors. This also was not disclosed. This lack of disclosure makes validation of the PMPRB study's results difficult. For the present article, we reverse-engineered the data from the figures provided in the PMPRB study and are confident that, in so doing, we introduced only an acceptable level of error such that the relative calculations of best fit, for example, are correct. Nevertheless, a court analyzing the PMPRB study cannot fairly evaluate its claims without access to the data or clear guidance on how to reconstruct the data from publicly available sources.

2.6 Peer Review

The PMPRB study did not undergo any form of external peer review. Although the PMPRB study states that "the NPDUIS Steering Committee [provided] ... expert oversight and guidance in the preparation of" the PMPRB study,⁵¹ this does not

⁵¹ *Ibid.* at i.

amount to peer review in which experts in both the subject matter under investigation and the methodology used would be called on to critically assess that methodology and the conclusions drawn. As the discussion of the PMPRB study's methodology and application of that methodology provided above makes clear, the PMPRB study would not have successfully passed such a review.

3.0 A PRACTICE FOR THE ADMISSION OF SCIENTIFIC STUDIES

The *Mohan*, *J.-J.-L.*, and *Trochym* trilogy invest the judge with the obligation to critically assess purported scientific studies before admitting them into evidence whether at the behest of a party or by judicial notice. Beyond noting general factors to assess, however, neither the trilogy nor subsequent law provides clear guidance to judges on the process through which to make this critical assessment. The discussion above concerning the introduction of the PMPRB study into evidence for the purposes of calculating damages provides a basis for developing such guidance. The remainder of this article suggests a stepped approach to determining whether a purported scientific study is even “good enough to be wrong.”

When a scientific paper is presented to a court, whether at the behest of the parties or through the court's own research, the first step is to examine the limitations placed on the paper's conclusions by the authors themselves. Such limitations are normally required to pass external peer review, because reviewers will be sensitive to authors overstating the import of their conclusions. Where, as in the PMPRB study, the limitations make clear that the scientific paper was not aimed at answering the question that happens to be before the court, the paper should not be admitted. Although parties may call experts to extend the learning from the paper, it is that expert's report—with its own stated limitations—that ought to be admitted (assuming it also satisfies other rules of evidence), not the original paper. This is necessary to ensure that the person with expertise in the field takes responsibility and is ready to be cross-examined on his or her opinion. The court should not be put in the position of relying on the opinion of the original author who is shielded from cross-examination.

Should it be determined that the paper is being introduced for a factual conclusion falling within the authors' intended limits, the judge must assess whether the methodology used in the paper follows well-established practice in the field or whether it proposes something new. “Not all scientific evidence, or evidence that results from the use of a scientific technique, must be screened before being introduced into evidence. In some cases, the science in question is so well established that judges can rely on the fact that the admissibility of evidence based on it has been clearly recognized by the courts in the past.”⁵² Thus, studies employing well-established methodologies used in well-established ways can be admitted.

⁵² *Trochym*, *supra* note 6 at para 31.

Where the methodology deviates from ordinary practice, however, the judge ought to examine the reasons for that deviation. Simply because the methodology used is different does not mean that the study constitutes bad science. However, the onus is on the authors to explain both how their methodology differs from established methods and to justify the deviation. The failure either to acknowledge that ordinary practice is not being followed or to justify that deviation should alert the judge that the methodology used does not constitute sound science. Without further justification through expert evidence presented to the court, the judge should be cautious about admitting the paper into evidence.

Where the methodology differs from ordinary practice, but the authors provide a justification for that deviation, the judge must next determine whether the justification provided is sound. There are two ways in which a judge can make this determination: (1) the paper and its methodology have undergone independent, external peer review or (2) an expert in the underlying methodologies provides evidence that the methodology is sound. In the latter case, the expert will have to be called by the party seeking to have the paper admitted. The expert ought to be able to speak to the methodology itself (as opposed to the conclusions reached) and could be either one of the authors of the report or a third party with substantial credentials in the methodologies used.

It is important to elaborate on the distinction made above between an expert on the methodology used and an expert in the substantive area covered by the paper. The expert in methodology will understand the forms of bias that different methodological approaches bring to a study and ways to address or limit those. This understanding is critical to the ability of the judge to assess whether the paper constitutes good science. On the other hand, an expert in the substantive field—for example, economics or management—may only be able to assess the reasonableness of the paper's results and not necessarily whether those results were reached using a valid methodology. Admissibility ought to be determined on the basis of methodology, not plausibility of results. The latter goes to weight.

In order for an expert to validate the methodology, the paper's authors will need to provide access to the data used and the set of criteria used in implementing the methodology. Only with this information will the expert be able to assess the biases and errors inherent in the methodology. (We note that all methodologies bring with them some level of error and bias. What is important is to understand those errors and biases in interpreting the results.) Although in the ideal case, the data used in the paper will be publicly available, this is often not practicable because the data may be subject to contractual limitations or be secret. In these cases, the party attempting to admit the paper into evidence ought to ensure that it makes the data available to the court, perhaps subject to an order of confidentiality, or that its expert is able to recreate the data from publicly available sources using criteria clearly set out by the paper's authors. Without access to the data, it may be impossible to validate the paper's methodologies and conclusions.

4.0 CONCLUSIONS

In summary, this article proposes that admissibility of purported scientific papers be evaluated in accordance with the following guidelines:

1. Does the use of the paper fall within the uses contemplated by the paper's authors—that is, have the authors noted any limitation or scope restriction on the use of their results? If the use does not fall within the contemplated uses, the paper ought not be admitted.
2. If the use of the paper does fall within the uses contemplated by the paper's authors, does the paper's methodology deviate from standard practice? The paper's authors would be expected to clearly state the methodology used, if and how it departs from standard practice, and the criteria used to apply that methodology. If the paper does not provide this disclosure (as will often be the case if it is not peer reviewed), the admissibility of the paper should attract great scrutiny. If the paper falls within standard practice, then it is likely to be at least good enough to be wrong. The determination whether the conclusions reached are correct goes to the question of weight rather than admissibility, a question that is outside the scope of the present analysis.
3. If the paper's methodology deviates from standard practice, the judge should ask whether it was peer reviewed by external and independent experts in both the substantive area covered as well as in methodology. Such a review can be assumed for a high-ranking journal. Otherwise, the party introducing the paper ought to demonstrate that a sufficient peer review was conducted on the substance and the methodology used. If the paper was sufficiently peer reviewed, it ought to be admitted into evidence.
4. If the paper was not sufficiently peer reviewed, was the court presented with evidence from an expert in the underlying methodology that the methodology used was sound? Was this expert provided with the data used or with clear instructions on how to derive the data from publicly available sources? If the court is persuaded after hearing this evidence that, on a balance of probabilities, the methodology used was sound, the paper ought to be admitted.

If this methodology were to be applied to the PMPRB study, one would quickly conclude that the study ought not to be admitted into evidence. First, it fails to satisfy the first test because the use being made of the PMPRB study—to determine utilization in the entire pharmaceutical market—clearly falls outside the limitations provided by the PMPRB study's authors. Second, the PMPRB study's methodology was not adequately described nor were deviations from standard practice explained. Because the PMPRB study was not peer reviewed, the party seeking to admit the PMPRB study would have the obligation of calling an expert in methodology to validate that methodology. Because the PMPRB study did not provide the data on which it relied, the party seeking admission would have to make the data accessible to the court.

The *Mohan*, *J-L.J.*, and *Trochym* trilogy places the responsibility on the judge to keep bad science out of the courtroom, but provides little practical guidance on how to do so. Through the examination of one social science report, this article has illustrated the types of methodological problems that render a study not sufficiently reliable even to be admitted into evidence. By carefully investigating the methodology used and its justifications, a judge can successfully insulate the judicial process from the effects of science that is not even “good enough to be wrong.”