

**ANALYSIS OF THE EXISTING CONSUMER
RESEARCH RELEVANT TO BANS ON THE
DISPLAY OF TOBACCO PRODUCTS IN
RETAIL OUTLETS**

Professor David Midgley

23 June 2016

CONTENTS

CLAUSE	PAGE
1. INTRODUCTION	1
2. EVALUATION CRITERIA.....	7
3. SUMMARY OF FINDINGS.....	14
ANNEX ONE – DETAILED ANALYSES OF RESEARCH CONSIDERED IN THIS REPORT	26
ANNEX TWO - ADDITIONAL MATERIALS CONSIDERED IN PREPARING THIS REPORT.....	120
ANNEX THREE – QUALIFICATIONS, RESUME AND SELECTED PUBLICATIONS	123

1. INTRODUCTION

Background

1. I am a Professor of Marketing at INSEAD in France. I have previously held visiting positions at the Anderson School, University of California, Los Angeles and, from 1987 to 1998, I was Foundation Chair of Marketing at the Australian Graduate School of Management, University of New South Wales, Sydney. From 2001 to 2005, I was elected Area Head for the Marketing Area at INSEAD.
2. Annex Three to this report sets out in detail my professional qualifications, my current resume and a list of sample publications that I have written. However, in summary terms and amongst other areas of expertise, I have authored over 120 publications, including papers in leading consumer research and marketing journals such as the *Journal of Consumer Research*, the *Journal of Marketing Research*, the *Journal of Marketing* and *Marketing Science*. Other scholars have cited my work over 5,000 times. I have also acted as peer reviewer for papers submitted by other researchers to journals such as these for over three decades and I have served on the editorial boards of the *Journal of Consumer Research* and the *International Journal of Research in Marketing*. In the context of this report, I have also worked on the topic of consumer purchasing of non-durable goods, from a consulting as well as academic perspective, and in executive education. In summary, my expertise centers on understanding how consumers react to marketing activities, which is also the subject of all the research studies I have examined in preparing this report.
3. I have been asked to prepare this report for Japan Tobacco International (“JTI”), as I describe in further detail below. I would note at the outset that the views in this report are my own, and that I have not previously been involved in any research relevant to tobacco. In addition, I have not met with any company executive or other employee of JTI.
4. The background to this report is legislation that bans the display of tobacco products in retail outlets (“the Display Ban”). I understand that Display Bans are in effect in a number of countries, including Canada, Croatia, Ireland, Iceland, Norway, Russia, Thailand, Finland, Australia, New Zealand, and the United Kingdom.¹
5. I have reviewed the following two reports prepared by Dr Warren Keegan (“the Keegan Reports”) for JTI which review, amongst other things, consumer

¹ I understand that, in certain of these countries, there are circumstances where a consumer may still be able to see tobacco products at point-of-sale prior to purchasing. For example, I understand that, in the United Kingdom, legislation permits an adult consumer to request to see a tobacco product prior to purchase.

research evidence² publicly available at the time of the Keegan Reports and relied on in support of a Display Ban:

- (a) “Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control”, dated 2 September 2008; and
 - (b) “Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control – a Supplemental Report”, dated 19 June 2009.
6. I have been asked to identify and then review the consumer research which has been made publicly available since the Keegan Reports and is relevant to, and which seeks to support the introduction of, Display Bans. Having done so, I have been asked to provide my expert opinion on the extent to which that research provides reliable evidence that Display Bans would be effective in achieving the public policy goals (identified by various regulators³) of changing actual smoking behavior, namely in:
- (a) Reducing smoking initiation among minors; and
 - (b) Increasing smoking cessation (also known as ‘quitting’) among minors and adults.
7. I have identified and reviewed research that is not previously discussed in the Keegan Reports, and notably research published between 2010 and 2016. Since social science builds on the thinking and findings of prior work, and seeks to continually improve methodologies, these studies should, in principle, provide the best evidence available in relation to Display Bans at the time of writing.

² As I discuss in more detail at paragraph 64 below, I use the terminology “consumer research” as a general heading to encapsulate the three main types of research that have been conducted in the field of Display Bans (namely one-time surveys, longitudinal surveys and experiments).

³ See, for example: (i) the Impact Assessment accompanying the New Zealand Smoke-Free Environments (Controls and Enforcement) Amendment Act 2011, available at: <http://www.health.govt.nz/about-ministry/legislation-and-regulation/regulatory-impact-statements/better-retail-controls-tobacco-smoke-free-environments-controls-and-enforcement-amendment-bill>); (ii) the Explanatory Memorandum to the Tobacco Advertising and Promotion (Display) (England) Regulations 2010 (http://www.legislation.gov.uk/uksi/2010/445/pdfs/uksiem_20100445_en.pdf); and (iii) the report “Australia: the healthiest country by 2020 - Technical Report 2” prepared by the Tobacco Working Group for the National Preventative Health Taskforce ([http://www.preventativehealth.org.au/internet/preventativehealth/publishing.nsf/content/96cac56d5328e3d0ca2574dd0081e5c0/\\$file/tobacco-jul09.pdf](http://www.preventativehealth.org.au/internet/preventativehealth/publishing.nsf/content/96cac56d5328e3d0ca2574dd0081e5c0/$file/tobacco-jul09.pdf)).

Consumer research considered in this report

8. My own process of identifying such publicly available research was comprehensive and accords with the type of searches that I consider appropriate to substantiate the views I set out in this report. By way of example, this process of research identification consisted of analyzing various journal publications, searching relevant titles in those publications, and reviewing various policy documents (including those cited in Annex Two). Therefore, whilst it may be the case that I have not identified every piece of research which some might consider in some way relevant to the issues considered in this report, I am of the expert opinion that I have been able to review a sufficient volume of materials to substantiate the views I set out in this report. Given the nature of the exercise I have conducted (an analysis of the research in support of the introduction of Display Bans), it would not be necessary nor relevant for me to have reviewed any JTI internal documents relating to the display of tobacco products (even if such documents do in fact exist).
9. Within the consumer research I have reviewed in preparing this report, there are a number of studies that purport to be of relevance to the issue of Display Bans and which seek to support the introduction of such bans. In reaching my conclusions, I focus on studies that meet the following criteria, for the reasons explained below:⁴
 - (a) My evaluation focuses solely on studies relating to the *in-store display* of tobacco products and not to the advertising or promotion of tobacco products in general. My reasoning for this focus is that if one is to assess the likely effects of a Display Ban on smoking initiation or cessation, it is necessary to have studies which separate the effects of in-store display from the forms of advertising and promotion that exist, both generally and specifically within the retail store environment (to the extent permissible). In-store displays have their own characteristics, which differ from advertising and promotion. Understanding the effects of in-store displays therefore requires that we can identify and measure these effects separately from other forms of advertising and promotion, as well as any other factors in the general or store environment that may influence consumer decisions. I note a similar point was made in the Keegan Reports.
 - (b) I have been asked to look at the evidence that a Display Ban would *change the behavior of consumers*, namely in reducing smoking initiation and/or increasing smoking cessation. This is an important distinction to make as many social scientists have pointed out that what people say they do and how they actually behave are not necessarily the same thing. There are many biases and errors that can distort self-

⁴ A list of the additional materials that do not meet the criteria I discuss in paragraph 9 but which I have reviewed in preparing this report is set out at Annex Two. In paragraph 10 below I also provide a table summarizing my reasons for excluding them from my main review.

report measures (Cook & Campbell, 1979; Schacter, 1999).⁵ Equally what people say they intend to do and what they eventually do are not necessarily the same either. There may be a significant gap between intentions and behavior due to many potential inhibitors (Sheppard, Hartwick and Warshaw, 1988).⁶ And this is even more the case when the gap is between intentions and changes to the habitual behavior of individuals (Webb & Sheeran, 2006).⁷ I have therefore sought to identify those studies that seek to measure changes in consumer behavior rather than those that solely report consumer opinion, perceptions or intentions.

I will return to this point in Section 2, where I set out my evaluation criteria, as this turns out not to be an easy goal to achieve. Most of the studies relevant to a Display Ban rely on the *self-report* of consumers, and only a few studies measure observed behavior in the sense experimental social scientists or marketing data analysts would understand. Again, I note a similar point has been made in the Keegan Reports.

- (c) I have focused on studies published in social science journals where the researchers collect or obtain *primary empirical data* on the behavior of individual consumers and analyze these data themselves in order to reach their conclusions. Such studies can, potentially, provide direct evidence in relation to a Display Ban, and because of the requirements of academic journals, should provide sufficient details on their methodology and data to allow an informed evaluation of the evidence they present. While standards vary between journals, most encourage authors to explain and justify their choice of methodology and to discuss any potential limitations to their data, measures and conclusions.
- (d) I therefore exclude published papers whose aim is to review the existing research literature, and whose discussion of any one study is therefore necessarily summarized, or those studies that solely rely on qualitative or group methodologies (such as focus groups) that are not designed to test hypotheses or provide data that can be generalized to a broader population. I also exclude discussion and/or working papers that are not published in scientific journals.

10. The 17 studies that meet my criteria for inclusion are discussed in detail at Annex One. I have included at the introduction of Annex One a summary table

⁵ Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues*. Boston, MA: Houghton Mifflin Company. Schacter, D. L. (1999). The seven sins of memory: insights from psychology and cognitive neuroscience. *American Psychology*, 54, 182-203.

⁶ Sheppard, Blair H., Jon Hartwick, and Paul R. Warshaw (1988), "The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research," *Journal of Consumer Research*, 15 (3), 325-43.

⁷ Webb, T.L. & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of experimental evidence. *Psychological Bulletin*, 132(2), 249-268.

listing the individual studies and the outcomes each set out to measure (as they relate to the objectives listed in paragraph 6 (a) and (b) above), together with a summary of my views on these results.

11. To assist the reader, I set out below my reasoning for excluding the additional materials listed in Annex Two from a detailed discussion in this report.

Additional material	Content	Reason (s) why it does not meet my criteria for detailed discussion in the report
ASH, 2013	Briefing document on display bans	Discussion paper/no primary empirical data generated.
Basham, 2010	Discussion paper on the effects of Canada's display ban.	Discussion paper/no primary empirical data generated.
Basham & Luik, 2011	Examination of the effects of display bans in four countries.	Literature review/no primary empirical data generated.
Berman et al. 2013	Review of tobacco product display restrictions	Discussion paper/literature review/no primary empirical data generated.
Cancer Research UK	Response to Basham, 2010.	Discussion paper/no primary empirical data generated.
Carter et al. 2009	Effect of pack displays on unplanned purchases	Reviewed in the Keegan Reports.
Clattenberg et al. 2013	Unplanned purchases and POS advertising	Small sample of unplanned purchases. Study does not separate the effects of tobacco displays at the POS from forms of retail advertising and promotion permissible in the USA.
Cohen et al. 2011	Study of retail compliance	Study does not consider data in respect of consumer behavior.
Curry et al. 2014	Historical analysis of a community display ban in Haverstraw, NY.	Study does not consider data in respect of consumer behavior.
Feighery et al. 2006	Evaluation of measures of exposure to in-store	Methodological study. Study does not separate the

	advertising	effects of tobacco displays at the POS from forms of retail advertising and promotion permissible in the USA.
Germain et al. 2010	Study of smoker sensitivity to tobacco displays and quitting.	Small sample of quitters, main focus is smoker sensitivity which is not smoking initiation or cessation behavior.
Hastings et al. 2008	Report on POS display of tobacco products	Discussion paper/survey research component subsequently published as MacKintosh et al., 2012.
Hoek et al. 2010	Study of 20 smokers who had attempted to quit.	Qualitative data only.
Jaine et al. 2014	Study of retailer attitudes to bans and other tobacco control measures	Study does not consider data in respect of consumer behavior.
Johns et al. 2013	Study of adolescent exposure to tobacco retail outlets	No primary empirical data relevant to Display Bans generated.
Li et al. 2015	Study of the impact of the display ban in Thailand using the ITC Southeast Asia Survey.	No data on smoking initiation or cessation behavior. These measures are not included in the Southeast Asia survey.
MacGregor, 2008	Consultant study of how tobacco products are displayed in stores	Study does not consider data in respect of consumer behavior.
Mackintosh et al. 2012	Study of association between POS displays and youth smoking susceptibility.	No data on smoking initiation or cessation behavior.
Monshouwer et al. 2014	Review of the literature on POS displays.	Discussion paper/literature review/ no primary empirical data generated.
Paynter & Edwards, 2009	Review of the literature on POS tobacco promotion	Literature review/ no primary empirical data generated.
Pereira & Veludo-de-	Study of retail POS formats	No primary empirical data on consumer behavior

Oliveira, 2014	and retailer opinions.	generated.
Robertson et al. 2014	Review of the literature on POS promotion and smoking	Literature review/ no primary empirical data generated.
Schmitt et al. 2015	Study of public support for display bans	Qualitative data. No data on smoking initiation or cessation behavior.
Thomson et al. 2008	Review of the evidence for and against display bans, interviews with ex-smokers, smokers and retailers.	Qualitative data. No data on smoking initiation or cessation behavior.
Wakefield et al. 2006.	Experimental study of the effects of POS advertising and pack displays on school children.	Reviewed in the Keegan Reports.
Wakefield et al. 2007.	Study of perceived effects of removing cigarettes on impulse purchase.	Reviewed in the Keegan Reports.
Wakefield et al. 2012	Study of brand placement on price boards using a retail audit.	Study does not consider data in respect of consumer behavior.

12. Finally, in the course of my own research, I have also identified and considered publicly available econometric studies that analyze official government data in the context of Display Bans (see paragraph 62 below).

Summary conclusion

13. In summary, my view is that, as a body of knowledge, the studies I have reviewed in this report do not provide reliable evidence to suggest that a Display Ban would be effective in achieving the policy objectives of reducing smoking initiation in minors or encouraging smoking cessation in minors and adults. I discuss my reasons for reaching this opinion in detail in Sections 2 and 3 below.

2. EVALUATION CRITERIA

Introduction

14. In this section, I set out the criteria by which I evaluate whether the studies I have considered in this report provide reliable evidence that a Display Ban would reduce smoking initiation among minors or increase smoking cessation among minors and adults.
15. I discuss my evaluation criteria in detail below but, in summary, these criteria are as follows.

- (a) Whether or not the study builds on a “middle-range theory” and derives hypotheses from this theory;
 - (b) The appropriateness of the research design given the hypotheses to be tested;
 - (c) The representativeness of the sample of individuals obtained;
 - (d) The validity, reliability and lack of bias in the measures employed in data collection instruments; and
 - (e) The appropriateness of the analysis methods used, given the hypotheses to be tested and the measures employed.
16. In formulating these criteria, I have drawn primarily on my training and extensive experience as a social scientist, particularly in reviewing articles submitted to the leading consumer research and marketing journals in my fields of expertise. I also reviewed the evaluation criteria set out in the Keegan Reports. While there is considerable overlap between my views and those of Dr Keegan, particularly on methodological and measurement issues, I have set out further additional criteria. This is primarily because while the journals I act as a reviewer for are equally cognizant of the difficulties and limitations of social science methodologies and measures, they also place a high value on the systematic accumulation and synthesis of knowledge into workable theory. Thus, I complement Dr Keegan’s emphasis on methodology with an additional emphasis on the value of grounded theory.
17. I will now discuss this point in depth as I consider it to be of overall importance to my evaluation of the reliability of the evidence provided by the studies I review in this report. I will then explain the remaining criteria.

Theory and hypotheses

18. Good social science advances the same way as other sciences. That is, researchers are expected to know what work has gone before and to build on this, both in terms of extending or refuting theory and in improving their research designs, measures and data collection methodologies from what has gone before.
19. Over time, specific domains within a field develop a theoretical framework about what drives their phenomenon of interest. This is sometimes called a middle-range theory because it does not attempt to explain everything (e.g. what makes people happy in life), rather it attempts to explain phenomena of a more limited scope (e.g. is job satisfaction related to happiness).
20. In general, a strong theory builds both on conceptual work by researchers, but also on a body of empirical findings from a number of related studies. These studies confirm or refute previous results, identify the boundaries of where the theory works and where it does not, and suggest what should be done next, but they all derive from the theory as an encapsulation of what the field knows about the domain at that point in time.

21. Because of this accumulation of knowledge, researchers are then able to derive and test specific (“sharp”) hypotheses in subsequent studies. For example, if one considers the leading consumer research journals (e.g. the *Journal of Consumer Research*), one will see published studies begin by reviewing prior theory and findings and then derive formal hypotheses from these, essentially defining the mechanisms by which certain events or factors (e.g. beliefs, motivations, intentions) will lead to certain outcomes (e.g. specific behaviors).
22. However, I was not able to apply this criterion to the studies I have considered in this report as few, if any, contain any theory or set up formal hypotheses. The typical study on Display Bans cites the objectives and empirical results from prior work, and then moves directly to a discussion of methodology. A few studies may briefly mention potential mechanisms for the way a Display Ban might work (often in one or two sentences), but none of these studies advances what I would call a middle-range theory for the effects of a Display Ban. This has three unfortunate consequences:
 - (a) First, we are left with an event, a Display Ban, purporting to be connected to a specific behavior (e.g. smoking initiation) but without any real explanation or formal hypotheses about how the removal of a display will change the way consumers think or behave, or indeed how the intermediate steps between event and purchase behavior may be influenced by the absence of a display. Such intermediate steps or factors are typically the core of any middle-range theory. For one small example, will the absence of a display of tobacco products mean consumers have less information at point-of-sale (“POS”) or will they simply access the needed information from another source?
 - (b) Second, because of the lack of theory, there are no agreed definitions for the factors that may influence behavior. Neither have the authors of the studies developed standardized, valid and reliable measures for these factors. The authors often have different definitions and measures and this creates a barrier to the accumulation of knowledge.
 - (c) Third, because there is no theory, we are also less well placed to understand the potentially unobserved factors, and methodological or measurement limitations, that may confound the results of the studies and that may lead us to mistakenly attribute their effects to a Display Ban when in fact they are due to something else. The more recent studies on Display Bans mention possible confounds and discuss measurement problems. However, lacking a theory, these discussions tend not to be systematic.

Overall, as far as theory and hypotheses are concerned, social science studies relating to a Display Ban do not resemble those in consumer research, social psychology or other major social science disciplines, in form or substance.
23. The other fact I observe from analyzing the consumer research in this report is that most studies here do not draw on the work of other relevant disciplines. This is an important oversight because there is a large body of scientific knowledge, and several middle range theories, relating to consumer

purchasing behavior in particular and marketing, advertising and promotion in general.

24. For example, the *Journal of Consumer Research* dates from 1974 and the *Journal of Consumer Psychology* from 1992. Both report large bodies of work on key aspects of consumer behavior, much of which is potentially of relevance to the studies considered in this report. There is also relevant work in many of the broader social psychology and psychology journals, as well as general marketing, advertising and promotion topics covered in the *Journal of Marketing* or the *Journal of Marketing Research*. Incorporating the learning from these disciplines and journals would have improved the designs, methodologies and reliability of the evidence gathered on the effects of a Display Ban. In particular, I note that much of the work on visual displays, purchase triggers and the like in consumer research generally is conducted through controlled laboratory experiments. With the notable exception of Kim et al. (2013, 2014) and Nonnemaker et al. (2016), such controlled experiments are missing from the field of Display Ban studies, limiting our overall understanding of the phenomena in question.

Appropriateness of the research design given the hypotheses to be tested

25. Any research design should accurately measure the factors that the researcher sets out as important to understanding the phenomenon in question. For example, if factor Y is important to your explanation of the phenomenon, but you actually measure factor X in your data collection, you are not testing your explanation as set out. In addition, and following the philosopher of science, Karl Popper,⁸ the research design should be capable of refuting (“falsifying”) any explanatory hypothesis the researcher put forward. Put another way, the research design has to be capable of showing the researcher’s prior ideas are wrong, otherwise there is no strong test of these ideas.
26. Another major issue with and source of confusion arising from the studies relied upon in support of Display Bans is the attribution of causality, i.e. whether factor Y causes factor X. A common mistake is to attribute causation to a study design that can only identify correlation, i.e. factor Y and factor X appear to associate with each other in the researcher’s data. To move from an attribution of correlation to one of causation often requires an understanding of timing, i.e. if we know Y happened before X, and we have an understanding of the mechanism linking the two factors, then it is plausible that Y causes X. This is an issue for many social sciences as the typical study is conducted at one point in time and therefore can only show correlation, not causation.
27. Here, most of the studies I have reviewed in this report are stronger than the typical social science study, since they use longitudinal designs with repeated surveys that, in principle, might be better placed to attribute causality. This is the strength of the body of research on Display Bans. In many disciplines, longitudinal designs, while often called for, are basically considered too expensive, too time-consuming or too difficult to conduct.

⁸ Karl Popper, *The Logic of Scientific Discovery*, London: Routledge, 2nd Edition, 2002.

28. On the last point, it is true that longitudinal field surveys are subject to possible confounding factors (e.g. events happening in the environment at the same time that also affect the phenomenon of interest but which are not measured by the researcher). In that respect, laboratory experiments can allow better control but sometimes at the cost of realism, a point I will return to later (see paragraph 274 in particular).

Representativeness of the sample obtained

29. In general, for policy-related studies, one would like to generalize the results of the study to the population at large. For example, if a researcher studies a sample of one thousand 18 to 21 year old males, they would like to say the research findings obtained from this sample are representative of *all* 18 to 21 year old males. In an ideal world, statistical theory tells us that if we draw our sample at random from a list of the total population of 18 to 21 year old males then we can make that generalization (assuming, of course, that those we select will answer our survey).
30. In reality, the world of consumer surveys is never ideal. Sometimes researchers cannot obtain a list of the target population from which to sample, strict random sampling from this list is sometimes also not possible, and most often people are not at home when a researcher calls or, if at home, refuse to answer their survey. All of this puts limitations on the researcher's ability to generalize findings from the survey to the target population. And generalization is particularly important if the objective of the research is to make policy recommendations, like those studies I review here, many of which conclude with recommendations or suggestions on tobacco control policies.
31. A significant issue here is indeed the problem of non-response to surveys. If those who do not respond are a significant proportion of the target population, the researcher may run the risk that the people who answered the survey are somehow different to those who did not, potentially biasing the findings. The most obvious example here is that it is difficult to survey busy people, resulting in surveys being based on the opinions of those with more time on their hands. Those who run surveys regularly adopt various strategies to mitigate these problems (e.g. oversampling certain groups, repeat calling on those not home first time or reweighting the sample data to reflect the target population). However, all of these raise statistical and analytical issues, and place a strong onus on the researcher to demonstrate that representativeness can be claimed (e.g. by follow-up studies or comparisons with known data).
32. Those who run laboratory experiments have a different strategy to achieve randomness. Here participants in an experiment are randomly assigned to the various "treatment" conditions, often including a control group that do not experience the treatment. Thus, if a randomly assigned group of 18 to 21 year old males shows an effect as the result of the treatment they received, and the control group of similar 18 to 21 year old males does not, this is evidence for the existence of that effect.

Validity and reliability of measures

33. Put simply, in the social sciences, a “valid” measure is one that measures what you claim it does (and not something else) and a “reliable” measure would give you the same measured values if you repeated it a second time (for some factor that is hypothesized as stable). Various techniques have been developed for assessing validity and reliability, and some social science disciplines, including consumer research, go to considerable lengths to develop valid and reliable measures.
34. A related issue is bias in measurement. Many of the measures we obtain in social science are based on what people tell us and there are many ways their answers might be biased. First, a researcher needs to be careful that their questions do not lead the person to give us the answers we want to hear (i.e. questions should be neutral). Second, some topics are socially sensitive and so there is a risk the respondent will try to tailor their answer to what he or she believes is socially “correct.” Smoking is potentially one of these sensitive topics and so one needs to assess any potential for such “social desirability” biases. Third, asking people to recall events in their past, especially minor events like the purchase of a fast-moving consumer product, runs the risk of inaccurate or incorrect answers. How many of us can remember accurately what we bought in the supermarket one month ago?
35. Overall, when one evaluates a particular study, one needs to examine the possibility of these and other biases in the survey or laboratory questionnaire used, as well as question the validity and reliability of the measures used.

Appropriateness of analysis methods, given the type of data collected and hypotheses to be tested

36. This is a broad topic and I will only touch on a few key issues here.
37. First, social science data comes in many forms, from simple “yes/no” answers to batteries of questions measuring a specific factor or personality trait. The analytical techniques social scientists use are equally diverse and all come with requirements for certain types of data and also assumptions about the inferences that can be drawn from this data by the technique. Thus, it is necessary to ensure the analysis technique is appropriate to the type of data available.
38. Second, nearly all these techniques are statistically based and calculate the probability that a relationship between two factors may exist in the population at large (for example, the probability that a correlation or a coefficient in a regression model is not zero). Traditionally, researchers in the social sciences have applied tests on these probability levels at the 0.05 level (also called the 5 percent level). This means that if we conclude the analysis of our sample shows an effect at the 0.05 level there is only a 1 in 20 chance (1/0.05) that this effect does not exist in the population at large (or put another way we have a low risk that we are misled by the sampling error present in all random samples).

39. Recently, a strong argument has been put forward that tests at the 0.05 or even 0.01 level (1 in 100) are far too weak. Some statisticians (e.g. Johnson 2010) now argue that the *minimum* standard for accepting the existence of an effect should be 0.005 level (1 in 200), or 0.001 level (1 in 1000) for claims of highly significant findings, which I note would be more in line with the statistical standards of the physical sciences. The reasoning behind these raised requirements is that recent research, and newer statistical methods, demonstrate that researchers using weak tests of significance are partly responsible for the lack of reproducibility of scientific studies (to which I would add this applies particularly in the social sciences, where inability to replicate results from study to study has become a major topic of discussion recently).
40. *Statistical significance* (e.g. the probability that an effect exists, given sampling error) should not be confused with *effect size*. Effect size is the magnitude of the effect and thus its importance or meaningfulness to policy makers or the world at large. Statistical significance is simply our level of confidence that an effect exists; and thus a pre-requisite for any discussion of causality or confidence in our hypotheses. Effect size tells us whether any significant effect is of sufficient magnitude to warrant consideration in policy decisions.
41. It is possible for an effect to be statistically significant but of small magnitude, particularly with large samples. For example, suppose our experimental treatment group or post Display Ban survey showed a measure of 9.9 on some outcome and our control group or pre-ban survey a measure of 10. Given large enough samples, this difference could be statistically significant but observers might say “so what” because the difference between 10 and 9.9 was of no practical use. Indeed, the “so what” question is important in evaluating research with policy implications, where cost/benefit calculations are also of concern. To introduce a new policy, we not only require statistical significance, but also confidence the magnitude of the effect is sufficient to warrant the introduction of this policy.
42. In the studies I have reviewed in this report, the *odds ratio* is often used as the statistic to compare before and after the Display Ban outcomes or to assess the change in an outcome between different values of some independent factor (for example, are people who visit stores three times a week more likely to smoke than those who visit only once per week). The odds ratio simply denotes the change in the odds of the focal outcome due to the Display Ban or variations in the independent factor. In the context of a decrease in the outcome, odds ratios of 0.7, 0.4 and 0.24 would be regarded in the social science literature as small, medium and large effect sizes. In the context of an increase in the outcome, the equivalent ratios are 1.4, 2.5 and 4.2. I derive these ratios by applying Chinn’s (2000) formula for converting odds ratio to the best-known measure of effect size, Cohen’s *d*.⁹ I also note these ratios provide overly generous tests of effect size if we are considering an outcome

⁹ Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(3) 3127-3131.

that occurs for less than 10 percent of the population (Chen et al, 2010).¹⁰ This can be the case with some of the outcomes in the studies I have reviewed, particularly free recall questions or quit attempts. I should also note that, given the objectives for my report, I focus more on effect sizes for behavioral outcomes and less on effect sizes for opinion or intention outcomes.

3. SUMMARY OF FINDINGS

The studies in support of a Display Ban

43. I have considered those studies listed in Annexes One and Two of this report to determine whether they provide reliable evidence that Display Bans would be effective in achieving the public policy goals listed in paragraph 6 (a) and (b) above, namely to reduce smoking initiation in minors and increase smoking cessation in minors and adults.
44. As noted above, certain of the publicly available studies set out to study the behavioral effects of a Display Ban in real or simulated settings, and to draw conclusions about the likely effectiveness of such a ban in achieving policy goals. I discuss those studies in detail at Annex One. In summary, my view is that several of these studies do not directly measure the behavioral outcomes relevant to the public policy goals set out at paragraph 6 (a) and (b) above. I also consider that those remaining studies that do measure relevant outcomes either have a restricted view of the purchasing process or certain limitations in their methodology that make it difficult to conclude that the results can be reliably generalized. To be fair to the authors of those papers, I should point out that several make this same point themselves.
45. Indeed, the strongest of the studies that I have reviewed, the econometric work of Irvine and Nguyen (2014) and the experimental work initiated by Kim et al. (2013) illustrate the difficulties of generalization given our current state of knowledge and methodological development. They also set up somewhat of a paradox. The econometric work uses large samples of self-reported smoking behavior *over several years* and shows *no effects* of actual Display Bans on this behavior. This is also consistent with several other studies showing no effects on similar behaviors over time. In contrast, the experimental work shows *a significant effect* of POS displays in a virtual experiment involving *one simulated purchase* of cigarettes. Are the stark differences between these two results down to methodology (for example, self-reports may be inaccurate, or experiments may not represent real world behavior)? Or are they due to our lack of understanding of the tobacco consumer's purchasing process over time (for example, how does one act of behavior relate to longer-term consumption)? In my opinion, the current state of research on Display Bans does not allow us to answer questions like these and this is a major reason why I conclude there is no reliable evidence in support of such bans.

¹⁰ Chen, H., Cohen, P. & Chen, S. (2010). How big is a big odds ration? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics-Simulation and Computation*, 39, 860-864.

The research on Display Bans is innovative but small, and with significant limitations

46. There are several strong points in the studies I discuss in Annex One, notably the use of longitudinal methodologies and openness to different sources of data. Many of these studies also recognize at least some of the weaknesses in previous work, and several also attempt to address these weaknesses by improved methodology.
47. However, in the end, the number of studies which seek to understand whether people are likely to change their behavior as a result of a Display Ban is a small body of evidence by social science standards, especially given the differing objectives, methodologies and measures employed by these researchers. More critically, in my opinion, most of these studies have one or more significant limitations, such as to cast doubt on whether their conclusions are based on reliable evidence or are simply informed speculation. Informed speculation can be valuable and several of these studies provide good platforms for improved research in the future. However, informed speculation is not a substitute for reliable evidence.
48. Hence, I have set out below my overarching views of the limitations of these studies as a reliable body of evidence, as well as my conclusions on what this evidence does or does not allow us to conclude on whether a Display Ban is likely to result in the behavioral changes regulators seek to achieve.
49. When contrasted with the best social science, the most severe limitation is the lack of a middle-range theory. I have already mentioned that this leads to three undesirable consequences: (a) the lack of understanding/sharp hypotheses about the process by which various events and factors lead to purchase; (b) lack of agreement on definitions and measures; and (c) an ad-hoc understanding of factors that may potentially confound results. Having reviewed these studies, I can put some depth behind each of these points by suggesting what would have needed to be done to produce more reliable evidence than exists to date.

Lack of understanding of the purchasing process

50. I make three points here.
51. First, tobacco products are a frequently purchased consumer product, purchased and consumed daily or weekly, potentially in different work/life situations and possibly obtained from more than one type of retail outlet.¹¹ Yet, in many of these studies, the researchers essentially took a snapshot of a limited number of purchases, or in some cases just one purchase, and used this to infer behavior over the many purchases consumers will make over any

¹¹ See, for example: Dawes, J. (2014). Cigarette brand loyalty and purchase patterns: an examination using US consumer panel data. *Journal of Business Research* 67, 1933-1943. Gordon, B.R. & Sun, B. (2015). A dynamic model of rational addiction: evaluating cigarette taxes. *Marketing Science* 34(3):452-470.

reasonable period of time. If the researchers had a more systematic theory of tobacco product purchasing and consumption, and one recognizing the frequently bought nature of the product, they might have broadened their sample of the consumer's purchasing life to provide a more accurate and broadly-based picture. Irvine and Nguyen (2014) make a related point on consumer "inventory" effects. Consumer researchers have studied this topic extensively to understand whether an in-store promotion simply advances a purchase that the consumer would have made anyway or creates extra demand by bringing new customers into the product category or by encouraging extra consumption. Without a reasonable sample of the multiple purchases made by an individual consumer, it is therefore difficult to draw correct inferences about their "behavior" and/or the effects of promotions.

52. Second, different retail advertising and promotional media serve different marketing objectives and have different effects on consumers (e.g. in-store advertising, price tags, layouts, store personnel, etc.). These distinctions are largely lacking in the Display Ban studies. Indeed, some studies aggregate the effects of different media together, risking severe confounds in their work. There is one notable exception in the work of Kim et al. (2013 and 2014) and Nonnemaker et al. (2016), and more researchers should follow their lead. I recognize here that the store environment differs by countries and the regulations in force at the time of each study. In my review, and where the researchers themselves did not discuss this, I have attempted to ascertain what the store environment might have been. However, it would also be helpful if studies in this field reported this critical information in more detail.
53. Third, this lack of understanding of the purchase process might be acceptable when a field of study is new and has little work to draw on. However, in the case here, not only is this field over 15 years old, but there is also a substantive body of research and theory to draw on from other social sciences in general and consumer research and marketing in particular. My conclusion is that, to build reliable evidence on the effects of Display Bans, researchers in this field should borrow more from their peers in other fields.
54. In particular, it would be useful to develop a more holistic picture of this consumer purchasing process, including multiple instances of purchase, inventory depletion and replenishment patterns, brand switching and substitution behavior (particularly given the existence of electronic cigarettes), purchasing from multiple types of outlet in different social and work situations, and also including a more comprehensive and nuanced assessment of all the various influences and cues in the family, peer, social, social media, traditional media and store environments that may influence change in consumer purchasing patterns over the short, medium and longer term. This would parallel some of the developments in consumer behavior research and marketing over a similar time span.

Lack of agreement on definitions and measures

55. Here I would make two points.
56. First, most social sciences arrive at clear definitions of what it is they seek to measure and then develop valid and reliable measures to do so. This does not

seem to be the case in the studies I have reviewed in this report. Not only does this make it difficult to compare studies but also many “measurements” are derived from single item questions. For some purposes this may be appropriate but for others, particularly given the nature of the purchasing process I describe above, we would have needed a well-formulated set of questions to derive valid and reliable measures.

57. Second, only four of the studies I have reviewed are based on the *observation* of behavior; most rely on individuals’ self-report of their own behavior, which may be subject to inaccuracies or biases. For example, the work of Irvine and Nguyen (2014) focuses on behavior (e.g. smoking prevalence and smoking consumption), but the measures come from surveys where consumers self-report whether they smoke and how much. The former (whether I smoke at all) is likely to be much more accurate than the latter (how frequently I smoke) which could be subject to recall and social desirability biases. This is not to say studies like that of Irvine and Nguyen (2014) are not valuable, they are, but rather than to produce reliable evidence we would also have needed more studies using check-out scanner panel methodologies, laboratory experiments or other forms of observational data to complement the survey studies, and also to help us better understand accuracy and bias.
58. Building on the last point, I would have also expected to see more studies based on laboratory experiments, beyond the ones of Kim et al. (2013, 2014) and Nonnemaker et al. (2016). These would provide another good complement to the survey studies in that experiments typically provide better control of confounding factors (through random assignment of participants to treatment conditions and through the use of control groups) and therefore allow us more confidence that effects, if observed, exist. Behavior can also be observed in laboratory settings. The downside of experiments is that it is sometimes difficult to make the setting realistic (i.e. is a virtual store actually like a real store?), and more importantly, it is often difficult to make the behavior we observe in the experiment the same as the individual would display in real life. These difficulties can limit the generalizability of results to the population at large.
59. The table in Annex One summarizes what was measured in each of the 17 studies I have considered in detail in preparing this report and draws conclusions on how these measures relate to the objectives of my report as set out in paragraph 6 above. Only two of these 17 studies directly measure smoking initiation in minors as a behavior and these two studies, in fact, find no evidence for an association between exposure to POS displays and smoking initiation (Bogdanovica et al. 2015 and Kim et al. 2013). Moreover, none of these 17 studies directly measure smoking cessation in minors and adults. What they do measure is either factors potentially influencing initiation or cessation behaviors (e.g. perceptions or intentions) or other behaviors such as smoking prevalence and cigarette consumption. Perceptions and intentions are *potential precursors* to behavior, but unless one understands the chain of events and factors connecting the two, one cannot simply assume that any change in these precursors automatically implies changes in behavior. Indeed,

typically the correlations between precursors such as these and behavior are small in magnitude.

60. Equally, smoking prevalence and cigarette consumption are the net effects of initiation, continued consumption and cessation, so unless one has separate measures of these components, there are difficulties in estimating what any change in smoking prevalence implies for initiation or cessation. This lack of direct measures of the behaviors important to regulators places significant limitations on the ability to draw policy implications from the studies I have reviewed in this report, and it would be useful if future research in this field incorporated such measures.

Confounding factors

61. When one conducts an experiment or survey, one would like to believe the items we measure in our questionnaires cover all the factors relevant to the behavior we wish to explain. However, there may still be factors that we do not know about or omit to measure that may impact this behavior. Human behavior is complex and driven by many factors within the person and in the social environment. This can potentially lead us to mistakenly associate any change in behavior with what we measure when in fact it is due to something else we did not measure. To social scientists, these are known as “confounding factors”; to econometricians this is known as the “omitted variables” problem. The following are some ways to seek to minimize the impact of these confounding factors:

- (a) **Control groups.** One standard solution to the “confounding factors/omitted variables” problem is a control group not exposed to the treatment (here a Display Ban). If we have confidence that both the treatment group and the control group are exposed to the same confounding factors (even though we do not observe these) then any effect that we observe of the treatment is more likely to be a real one. Of course, outside the laboratory, control groups are not always easy to set up. This is why countries like Australia and Canada have been the more useful places to conduct Display Ban studies, and countries like Ireland or the Iceland have been less useful. In the former, the different states and provinces introduced Display Bans at different times and so there are natural control groups (states and provinces without bans at these times). In the latter, the Display Ban was introduced across the whole country at the same time, and so researchers are, in essence, forced to use “before” ban measures as a control for “after” ban measures. This is less robust as a methodology because the confounding factors may also have changed in the time between the two measurements.
- (b) **Control variables.** The other solution to this problem is to measure a broader set of variables that the researcher considers may have impact on the focal behavior. These are known as “control variables” and are not part of any theory the researcher may have, rather they account for factors outside the theory that may have impact on the behavior. In the analysis phase the impacts of these control variables are subtracted

from the observed changes in behavior, leaving the residual effects as a consequence of the theoretical factors—for example here the Display Ban.

62. **The importance of price changes as a possible confounding factor.** Several of the studies I have reviewed in this report use the control variable approach, typically controlling for demographics to separate the effect of a Display Ban from the effects due to these demographic factors. However, the three econometric studies I have reviewed (Lilico, 2009;¹² Padilla, 2010;¹³ and Irvine and Nguyen, 2014) show that there is potentially a major confound that many researchers in this field do not consider, namely the price of tobacco products (and the associated issues of individual income and the illegal trade in tobacco products). These three studies demonstrate the demand for cigarettes to be highly price sensitive. As the price of legal cigarettes goes up, demand for them goes down and sales of illegally traded cigarettes presumably go up (though reliable data on contraband is hard to obtain). As a consequence, any study that does not control for price changes during the time a Display Ban is introduced, which is all of the remaining studies, may have a major confound in their findings because they misattribute any observed fall in their measures to the Display Ban when in fact it is due to a price increase. It is striking that, in the studies conducted in Canada, once price is controlled for, there are no significant correlations between the introductions of Display Bans in the various provinces and changes in smoking prevalence or consumption.
63. **Media coverage as another confounding factor.** The other potential confounding factor I would raise, and to which no study I review here gives the serious consideration they should, concerns the media and social media coverage surrounding the introduction of a Display Ban (e.g. press, radio, television, Facebook). This coverage may well also have an effect on smoking behavior, possibly only a short-term effect, but one that should be controlled for in Display Ban studies, particularly those that take their measurements around the time of the introduction of a ban (e.g. Li et al. 2013).

Limitations of questionnaires and surveys

64. Here, I would identify two important limitations in most of the studies I have reviewed in this report. Firstly, the potential for social desirability biases given the sensitivity of smoking as topic for survey or other research with human subjects. Secondly, problems of non-response that limit the generalizability of findings to larger populations.
65. **Social desirability biases.** While other biases, such as recall bias, have been the topic of discussion in this field (e.g. Carter et al. 2015), social desirability biases, while sometimes mentioned in the discussion or limitations sections of these papers, do not receive the same depth of consideration and assessment as in other social sciences. For example, in cultural research, social desirability

¹² Lilico, A. (2009). The impacts of restrictions on the display of tobacco products. *Europe Economics*.

¹³ Jorge Padilla. (2010). The effectiveness of display bans: the case of Canada. *LECG*.

biases can be problematic and researchers need either to take steps to phrase questions to reduce this bias or undertake analyses to demonstrate that it is not a major limitation on their findings. For Display Ban studies, most countries have introduced other tobacco control measures in the past and there are often supporting anti-smoking campaigns in the advertising media as well as press coverage of smoking and health-related topics. Consumers are therefore aware that smoking is discouraged by their society and may face internal conflicts when asked about their own behavior toward this product in surveys or other forms of research. This can lead some individuals to give more socially acceptable but inaccurate answers about their behavior. Given the potential sensitivity of smoking as a topic, particularly in surveys of children and adolescents, I would have therefore expected more attention to social desirability and other biases in this research literature. This is less of an issue for laboratory experiments or surveys with control groups, provided we can assume these biases are the same for the treatment and control groups, but the majority of the studies I have examined do not have control groups and therefore need to consider the potential for such biases.

66. **Non-response bias.** For studies that seek to generalize their findings to a larger population, researchers need to take follow-up steps or conduct analyses to demonstrate that those that respond to the survey are no different to those that do not respond on dimensions that are relevant to the study objectives and/or relevant theory. Only if it can be shown that there are no significant differences between respondents and non-respondents is it then permissible to generalize findings to the population at large. On the other hand, if there are significant differences between the two groups then study findings can only be taken to apply to the particular sample obtained. The other, related, issue with non-response is the possibility it will lead to the confounding of effects. Burton et al. (2012) point out that if respondents self-select into a survey because of some unobserved but important factor that is linked to the behavior in question then the researcher may mistakenly attribute effects to factors in the survey when in reality behavior is driven by the unobserved factor. I raise these issues because in many of the studies I review here the authors take little or no consideration of any potential for non-response bias in their research. Yet many of these authors, at least implicitly but sometimes also explicitly, generalize their findings to make public policy recommendations for the population at large.
67. In conclusion, and as I discuss in detail throughout Annex 1, the studies I have reviewed in this report do not appear to take these steps to either reduce or assess these biases, and that places further limitations on their findings and conclusions.

Strengths and Limitations by Type of Research

68. The studies I have reviewed in this report cover a variety of different research approaches, so I consider it useful to outline the strengths and limitations of these studies according to their approach.

69. In general, there are three types of research in this field: (a) studies based on one-time surveys; (b) studies based on longitudinal surveys; and (c) laboratory experiments, and I discuss these individually below.

(a) **One-time surveys.** Since this type of study is designed to meet specific research objectives, they can contain multiple measures on topics of interest to the researchers. Like all surveys, this type of research potentially suffers from various biases and because they are conducted at one time, at best they can only establish correlations, not causality. Since they are typically individual academic projects they may or may not have adequate funding to obtain large sample sizes, so lack of adequate statistical power to detect effects can be an issue. However, here I would say that most of the studies I have reviewed have adequate sample sizes and several have large samples by social science standards.

(b) **Longitudinal surveys.** This is where the same surveys are repeated at different points in time to track behavior or establish trends. Several of the studies I discuss in this report which utilize this methodology seem to have capitalized on the availability of data from government or market research agency tracking surveys. These have the advantage of being well funded, with consequent large sample sizes, but researchers may have to accept the existing questions in these surveys, or at best be limited in the number of questions they can add to them. In principle, longitudinal surveys of this type could begin to address causality, however most of the research designs of the studies I have reviewed in this report do not permit this. The main reason for this latter limitation is that each survey uses a fresh sample of people, making it difficult to infer whether any observed changes are due to differences between the samples or real changes. There are few panel studies (that is, where the same individuals are interviewed at multiple points in time) or more complex longitudinal designs (which could also compensate for any additional biases introduced by repeatedly interviewing the same people).

Arguably one of the better designed and administered longitudinal survey is the Canadian Tobacco Use Monitoring Survey (“CTUMS”) on smoking prevalence in various age groups in Canada.¹⁴ These surveys were conducted monthly on large samples of people for over a decade (albeit I understand that it ended in 2012 and was replaced by the Canadian Tobacco, Alcohol and Drugs Survey¹⁵). The best analyses of the CTUMS data are provided by the three studies I have reviewed in this report which are conducted by econometricians (Irvine and Nguyen, 2014, and also Padilla, 2010¹⁶ and Lilico, 2009¹⁷). This is

¹⁴ See: http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/research-recherche/stat/_ctums-esutc_2012/ann_summary-sommaire-eng.php.

¹⁵ See: <http://healthycanadians.gc.ca/science-research-sciences-recherches/data-donnees/ctads-ectad/index-eng.php>.

¹⁶ Jorge Padilla. (2010). The effectiveness of display bans: the case of Canada. *LECG*.

because their field of research focuses on the substantive and methodological issues in analyzing complex mixtures of cross-sectional and time series data, and provides them with a number of analytical techniques and statistical tests with which to do so. In particular, econometrics stresses the importance of control groups and appropriate use of data and control variables to minimize the impact of omitted variables. There is also a tradition of trying a range of alternative analyses to establish whether any findings are artifacts of the researcher's choice of analysis technique or robust across alternative analyses. The other longitudinal survey studies I have reviewed in this report (e.g. Li et al. 2013) generally take a less sophisticated approach than the econometricians, using more basic "before/after" designs and simpler analysis techniques and tests.

That said, the econometricians appear to view the survey data as relatively accurate, or at least advance the same argument as in other studies cited here, namely that any survey biases or remaining confounds should be the same across provinces and time periods and therefore of less concern. While this is an entirely reasonable stance, I would have liked to see more work done to validate this conclusion. I note that both sets of researchers rather ignore the problem of non-response and generalizability.

- (c) **Experiments.** As noted above, there are only three experiments in the studies I have reviewed in this report, all of which use the same virtual store developed by the authors (Kim et al 2013, 2014) and Nonnemaker et al. (2016). Also, as noted above, these experiments have high internal validity (i.e. one is confident that the effects observed are real) but as is typically the case with experiments, there are questions of external validity (i.e. how realistic is the behavior observed in the experiment compared with behavior in the real world setting in which purchase decisions are actually made).

What studies have not been conducted that would significantly add to the reliability of the body of evidence?

70. In my opinion, the striking gap in this set of studies is the absence of studies that follow up on the longer-term effects (if any) of a Display Ban. As I noted in paragraph 4 above, Display Bans have been introduced in several countries and have now been in effect for several years (e.g. in Canada since 2004).
71. However, most of the studies related to Display Bans examine effects only as far as one year after the introduction of the ban (and some only at much shorter periods). So, in reality, we do not know whether there are any longer-term effects of a Display Ban. The best place to study this is likely Canada given the availability of the CTUMS data (until the CTUMS survey concluded in 2012). In Canada, smoking prevalence and consumption has been on a declining trend for some time, so one would need to extract this trend from the

¹⁷ Lilico, A. (2009). The impacts of restrictions on the display of tobacco products. *Europe Economics*

data, and then assess whether there were any longer-term changes in behavior due to the Display Ban or not.

72. In my opinion, completing such an analysis would have significantly added to our understanding of whether or not Display Bans are capable of achieving the public policy objectives of reducing smoking initiation among minors, and increasing cessation among minors and adults.

Conclusion

73. In discussing my overall conclusion, it is first useful to outline briefly the decision process by which an individual might chose to, for the first time, or continue to, purchase tobacco products. This process is linked to individuals' perceptions and beliefs about smoking, their motivation to either initiate, continue or cease smoking and then the actual behaviors themselves, namely smoking initiation, smoking consumption and smoking cessation.
74. I have been asked to consider the body of research on Display Bans as they relate to two of these behaviors, smoking initiation and smoking cessation, the former amongst minors and the latter amongst minors and adults. This is because it is these specific behaviors that are cited by those who seek to introduce Display Bans as being likely outcomes of such a ban.
75. What I find in the studies is that there is little research on the relationship between Display Bans and the two behaviors of actually starting and stopping to smoke; rather some of the research relates to earlier steps of the decision-making process, that is, perceptions, beliefs and intentions, and some relates to smoking prevalence or cigarette consumption, typically relying on self-reports of these behaviors.
76. The difficulty here is that what people think and what they actually do are not always the same thing, a problem that is compounded by the majority of the studies being based on individuals' self-report of their perceptions and beliefs and recall of their behaviors, or their opinions about what others might do or think. In the context of tobacco products, factors which might mean that existing smokers who indicate that they intend to quit, but do not then quit (or do not quit immediately) may include that (a) some smokers may feel that they ought to identify such an intention to the researcher (as they feel uncomfortable saying that they do not want to quit); or (b) "*many people who quit smoking relapse at some point*" and "*...it can take a few tries to quit smoking for good*".¹⁸ Either factor may mean that a "quit" question asked at one point in time is an inaccurate statement of the individual's real behavior.
77. The second point is that all steps of this decision-making process are subject to a large number of potential influences from the social environment surrounding the individual, for example social group norms and peer pressure, the availability to those intending to quit of alternative products (such as e-cigarettes or nicotine replacement products), public health campaigns

¹⁸ See <http://www.nhs.uk/Livewell/smoking/Pages/relapse.aspx>, a website of the UK National Health Service.

reminding people about the health risks of smoking and other regulatory initiatives. Price is also a major factor in any decision to purchase, and prices may change according to demand and supply, but also tax changes and the availability of illegal tobacco products (Irvine and Nguyen, 2014).

78. All of this makes survey research on tobacco product purchasing difficult, especially when the researcher wants to isolate the effects of one factor (in this case, a Display Ban) amongst many potential factors. I mention this point because I consider that, generally, the studies do not deal adequately with these potential confounding factors, and therefore risk attributing effects that they find to Display Bans when they may be due to something else in the environment.
79. With this backdrop, I now turn to my conclusion on whether the studies I have reviewed in this report demonstrate that a Display Ban is capable of changing actual smoking behavior as discussed in paragraph 6 (a) and (b) above. In short, the studies are not capable of reliably supporting a hypothesis that a Display Ban will have any impact on these behaviors. To demonstrate why this is the case, I will discuss, in brief, the results of four pieces of research that are most capable of answering this question, given their understanding of the individual consumer and their methodological rigor in isolating the effects of Display Bans from other factors. These are the study of Irvine and Nguyen (2014) and the three related studies of Kim et al. (2013, 2014) and Nonnemaker et al. (2016).
80. Irvine and Nguyen (2014) subject large samples of individual data from what is arguably one of the better designed and administered longitudinal survey on tobacco usage, the CTUMS data, to econometric analyses. In addition, the history of Display Bans in Canada, with different provinces adopting bans at different times, provides a natural experiment with the control groups that are not available to several other studies, but which are critical to isolating the effects of these bans amidst many other potentially confounding factors.
81. Irvine and Nguyen seek to measure smoking prevalence (which is a net measure of initiation, consumption and cessation), cigarette consumption, quit intentions and quit attempts before and after the introduction of the Display Ban. Despite using a differences-in-differences approach to control for unobserved confounding factors, employing a range of alternative analyses and having high statistical power through their large samples, Irvine and Nguyen find no significant statistical association between the introduction of a Display Ban and smoking prevalence, quit intentions or quit attempts (this being the case for both minors and adults). They do report some reduction in consumption for 15 to 24 year olds, but they note this is a statistically weak effect that only appears when they make a particular assumption about time trends in the data and does not appear with alternative assumptions. Therefore, overall this study would support a conclusion that a Display Ban is not likely to have any impact on smoking behavior in general.
82. Kim et al. (2013, 2014) and Nonnemaker et al. (2016) use a virtual store to conduct three related experiments, one on minors, the others on adults. Using an experimental approach allows the authors to test the effects they observe

through randomly assigning participants to treatment conditions and also to separate the effects of enclosed displays from in-store advertising in the first experiment and health warnings in the second. In the first experiment, the authors find no evidence of tobacco displays encouraging smoking initiation amongst minors. They do find enclosed displays reduce purchase attempts amongst minors compared to open displays. In the second and third experiments with adults, the authors find the same effect, namely a reduction in purchase attempts when displays are enclosed. In this regard, there is a question of how well the experimental task produces behavior that is similar to that of individuals shopping in real stores. That question is not yet answered. Indeed, the authors are careful and correct to state that their results derive from this virtual setting, and whether they translate to a real setting (that is, have external validity) is yet to be established. Overall, this stream of research does not, as yet, provide reliable evidence for or against the introduction of a Display Ban.

83. In my opinion, the other 13 studies are, given their design and methodology, not as capable at providing reliable evidence as the above, but it is interesting that two of three Display Ban studies that also measure behaviors (as opposed to perceptions or opinions) report no significant associations between the introduction of a Display Ban and these behaviors. These are McNeil et al. (2011) who demonstrate no statistical association between the introduction of a Display Ban in Ireland and smoking prevalence in the “short-term”; and Quinn et al. (2011) who show no significant association between the introduction of a Display Ban in Ireland and the sales of cigarettes, also in the “short-term”. In contrast, Dunlop et al. (2015) report a small (my interpretation of the effect size) “medium-term” association with current smoking levels. The relatively short time horizons of these studies (12, 12 and 24 months respectively) also lead directly to my next reservation on the reliability of the evidence provided.
84. An important question that is not answered by any of the studies is whether there are any potential longer-term effects of a Display Ban. There has been no follow-up research in countries where this might be possible and although I recognize that designing such a study is not a simple task, it would surely have been worthwhile to attempt to do so. The failure to do follow-up studies means we have no evidence on any longer-term effects of a Display Ban, should they exist.
85. In conclusion, I find that these studies do not provide reliable evidence supporting the policy objectives of reducing smoking initiation in minors or encouraging smoking cessation in minors and adults.

ANNEX ONE – DETAILED ANALYSES OF RESEARCH CONSIDERED IN THIS REPORT

In this Annex, I review the individual studies that meet the criteria outlined in paragraph 9 above against the evaluation criteria I set out in Section 2. To assist the reader, I first set out a summary table of these studies against the stated objectives of a Display Ban, discussed in paragraph 6(a) and (b) above. The studies are listed in alphabetical order of the first author.

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
Bogdanovica et al, (2015).	To investigate the association between the frequency of visiting shops and the development of susceptibility to smoking, or smoking uptake, in UK secondary school students.	<p>Study conducted before the Display Ban came into effect and therefore cannot directly assess the impact of such a ban.</p> <p><i>Smoking initiation in minors.</i> The study shows no significant association between smoking initiation (as measured by change in “smoking status” over a twelve month period between two surveys) and exposure to POS displays.</p> <p><i>Smoking cessation in minors and adults.</i> This study does not address adults. The authors do not report on the minors who ceased smoking between the two waves of surveys.</p>
Burton et al., (2012).	To assess the impact of retail displays on tobacco smoking and purchase by smokers and attempting quitters in NSW Australia.	Study conducted before the Display Ban came into effect and therefore cannot directly assess the impact of such a ban. The data in this study was collected just before NSW restricted tobacco displays solely to the POS (2008) and is therefore conducted in a more open promotional environment.

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
		<p><i>Smoking initiation in minors.</i> This study does not address minors.</p> <p><i>Smoking cessation in minors and adults.</i> For adults, this study mainly focuses on purchase and consumption, which are different behaviors to cessation. While “attempting quitters” are included in the sample, the authors do not analyze or draw conclusions on cessation. In addition, their measure of exposure is to all “retail displays”, which may include but are not the same as the POS displays subject to bans.</p>
Carter et al, (2015).	To assess the impact of a Display Ban in Western Australia on spontaneous purchase behaviors, defined as whether respondents intended to buy before entering the store or not.	<p><i>Smoking initiation in minors.</i> This study does not address minors or measure initiation, since those interviewed were adult smokers 18 years or more of age.</p> <p><i>Smoking cessation in minors and adults.</i> Reduction in spontaneous purchasing for adults reported in exit interviews, and correlated with less perceived influence of displays might potentially provide circumstantial evidence on this objective. However, smoking cessation was not specifically measured, we do not have a full picture of the consumer’s purchasing behavior, only one event, and the study has small sample sizes.</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
Dunlop et al., (2015).	To assess the medium term impact on youth smoking of the Display Bans introduced in NSW and Queensland in July 2010 and November 2011 respectively.	<p><i>Smoking initiation in minors.</i> While this study includes minors, and they are specifically identified in the analyses, smoking initiation as a specific behavior is not measured.</p> <p><i>Smoking cessation in minors and adults.</i> The post-Display Ban decreases in levels of brand awareness, normative beliefs and perceived impact of displays shown in this study are better thought of as factors potentially influencing behavior, rather than measures of actual smoking cessation. Smoking cessation itself is not measured and while the observed decreases in current smoking levels after the ban might provide indirect evidence that some individuals have quit smoking, the overall effect here is small in size.</p>
Haw et al., (2014).	To propose a protocol for examining the impact of the Display Ban in Scotland, including multi-modal, before and after design and mixed data collection methods.	This paper presents a proposal for a program of research and, at the time of writing, data had not been collected. Hence this study itself provides no evidence relevant to either of the objectives of my report. The study by van der Sluijs et al., (2016) reports analyses of the first data from this program and is discussed below.
Irvine and Nguyen, (2014).	To examine the impact of Display Bans by analyzing them as event studies. This methodology looks at the impact of an event—a Display Ban--on a time series of data on smoking prevalence and consumption in various provinces of Canada.	<i>Smoking initiation in minors.</i> This behavior is not directly addressed by this study. The study does include smoking prevalence but, as the authors point out, smoking prevalence is a net measure of initiation and quitting and thus not initiation itself. Furthermore, the study finds no significant association between the introduction of Display Ban

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
		<p>and smoking prevalence.</p> <p><i>Smoking cessation in minors and adults.</i> Quit intentions are one measure in this study but intentions are not the same as quitting itself. Further, the study finds no significant associations between the introduction of a Display Ban and quit intentions. The study also looks at quit attempts, which are more closely related to smoking cessation, but again finds no significant associations.</p> <p>What the study does report is limited support for some reduction in consumption amongst those in the 15 to 24 age groups, given certain assumptions about time trends in the data.</p>
Kim et al., (2013).	<p>To assess whether youth exposed to Display Bans in a virtual store are less likely to attempt to purchase and more likely perceive that it is difficult to purchase at that store.</p> <p>A secondary aim is to examine whether the presence of tobacco advertisements in the (virtual) store moderates those relationships.</p>	<p><i>Smoking initiation in minors.</i> This study focuses on 13 to 17 year olds and so addresses minors. The study also analyses “open to smoking” participants separately to those who are current smokers, thereby also addressing initiation. However, the most relevant analysis the authors provide shows no significant effect of enclosing the display on this “open to smoking” subgroup. Hence there is no evidence for initiation in minors.</p> <p><i>Smoking cessation in minors and adults.</i> Here the study provides experimental evidence that enclosing the display does reduce purchase attempts by minors. However, reducing purchase attempts during one store visit is not the same behavior as quitting to smoke,</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
		which is not specifically addressed by this study. The study does not address adults.
Kim et al., (2014).	To assess whether exposure of adults to enclosed tobacco display and a graphic health warning sign decreased urges to smoke and tobacco purchase attempts in a virtual store.	<p><i>Smoking initiation in minors.</i> This study does not consider minors, only adults.</p> <p><i>Smoking cessation in minors and adults.</i> This study focuses only on adults and does not specifically measure smoking cessation, rather it measures purchase attempts in a virtual store experiment. The experiment shows that purchase attempts in a virtual store decline when displays are enclosed. Whether a decline in purchase attempt leads to actual smoking cessation is not addressed by this study.</p>
Li et al., (2013).	To examine the immediate and long term consequences of PoS marketing. To examine the variability in PoS market restrictions across multiple countries and examine the effect of these varying restrictions on adult smokers' exposure to marketing and cigarette purchasing.	<p><i>Smoking initiation in minors.</i> This study does not specifically measure initiation and surveys adults, not minors. Hence it does not provide any evidence against this objective.</p> <p><i>Smoking cessation in minors and adults.</i> Smoking cessation is not specifically measured. Moreover, any circumstantial evidence on cessation that might be inferred from their purchase measure is confounded by the failure to separate the display of tobacco products from other in-store advertising and also doubt as to whether this item measures consumption or substitution.</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
Li and Walton, (2015).	To examine support for a Display Ban amongst smokers by looking at support for the Display Ban immediately before, immediately after and one year after the introduction of the Ban in New Zealand.	<p><i>Smoking initiation in minors.</i> This study focuses on adults, not minors.</p> <p><i>Smoking cessation in minors and adults.</i> This study does not measure smoking cessation or related factors.</p>
McNeil et al. (2011).	To evaluate the short-term impacts of the Display Ban in Ireland.	<p><i>Smoking initiation in minors.</i> This study does include minors but has no specific measure of smoking initiation as a concrete behavior. The measure included, smoking prevalence, is a net measure of initiation and quitting. Moreover, the study shows no statistical association between the introduction of the Display Ban and smoking prevalence amongst youth.</p> <p><i>Smoking cessation in minors and adults.</i> While there are measures of perception and opinion concerning smoking cessation, there is no measure of this as a specific behavior. Again, the study shows no statistical association between the introduction of the Display Ban and any fall in smoking prevalence.</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
Nonnemaker et al. (2016).	Examine the potential impact of a Display Ban and mandating plain packaging at the point-of-sale through an experiment.	<p><i>Smoking initiation in minors.</i> This study focuses on adults, not minors.</p> <p><i>Smoking cessation in minors and adults.</i> This study focuses only on adults and does not specifically measure smoking cessation, rather it measures purchase attempts in a virtual store experiment. The experiment shows that purchase attempts in a virtual store decline when displays are enclosed. Whether a decline in purchase attempts leads to smoking cessation is not addressed by this study.</p>
Paynter et al. (2009).	Examine the association between exposure to tobacco displays at the POS and teenager smoking and susceptibility to the uptake of smoking.	<p>Study conducted before the Display Ban came into effect and therefore cannot directly assess the impact of such a ban.</p> <p><i>Smoking initiation in minors.</i> The study addresses minors 14 and 15 years old but measures their “susceptibility to initiation” which is an intention and not the behavior of starting to smoke. The study shows an association between this intention and exposure to POS displays of small to medium magnitude.</p> <p><i>Smoking cessation in minors and adults.</i> The study does not address adults and the primary measure that might potentially relate to cessation in minors confuses those who have stopped to smoke with those who smoke infrequently.</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
Quinn et al. (2011).	Examine the impact of a Display Ban on cigarette sales through the use of statistical analysis techniques.	<p><i>Smoking initiation in minors.</i> This study does not address minors as such since their cigarette purchasing is not identifiable in the aggregate sales data it uses.</p> <p><i>Smoking cessation in minors and adults.</i> The study does not specifically address smoking cessation. What it does show is no significant correlation between the introduction of the Display Ban and the sales of cigarettes.</p>
Scheffels and Lavik, (2013).	To evaluate retailer compliance and consumer perceptions of and experiences with the Display Ban in Norway.	<p><i>Smoking initiation in minors.</i> I conclude this study does not provide reliable evidence about reducing smoking initiation in minors, since their measures of “<i>temptation at the POS display, purchase difficulty and brand choice</i>” are for individuals who already smoke.</p> <p><i>Smoking cessation in minors and adults.</i> The questions on the difficulty of purchase and brand choice might potentially provide some circumstantial evidence against this objective but the study results suggest little change in these responses either immediately or 10 months after the Display Ban.</p>
Spanopoulos et al. (2014)	Investigates the association between tobacco displays and brand communication at the POS and adolescent smoking behavior, and assesses the potential benefits likely to accrue when a Display Ban is introduced in the UK.	<p>Study conducted before the Display Ban came into effect and therefore cannot directly assess the impact of such a ban.</p> <p><i>Smoking initiation in minors.</i> The study shows no significant association between “ever-smoking” and noticing POS displays. The</p>

Name of study	The stated outcomes the study sought to measure	The results of the study as they relate to smoking behavior
		<p>“ever-smoking” measure aggregates together those individuals who are current smokers and those who have quit and hence does not provide a measure of initiation.</p> <p><i>Smoking cessation in minors and adults.</i> Smoking cessation is not specifically addressed by this study and as above “ever-smoking” aggregates current smokers and those who have quit. The study does not address adults.</p>
van der Sluijs et al. (2016)		<p>Study conducted before the Display Ban came into effect and therefore cannot directly assess the impact of such a ban.</p> <p><i>Smoking initiation in minors.</i> The study focuses on brand awareness, which is not a behavior as such. The study shows small (and possibly non-significant) associations between exposure to POS displays and brand awareness.</p> <p><i>Smoking cessation in minors and adults.</i> The study does not address smoking cessation or adults.</p>

Ilze Bogdanovica, Lisa Szatkowski, Ann McNeil, Dionysis Spanopoulos and John Britton. Exposure to point-of-sales displays and changes to susceptibility to smoking: findings from a cohort study of school children. *Addiction* 110.4 (2015): 693-702 (Bogdanovica et al., 2015).

Objectives, background and theory

86. The authors state their aims are to “*investigate the association between the frequency of visiting shops and the development of susceptibility to smoking, or smoking uptake*” (p. 693), in UK secondary school students.
87. They cite a number of studies from the prior literature, including the studies of Dunlop et al. (2015) and Kim et al. (2013, 2014) that I also review in this report. However, in their view, while this literature shows associations between point-of-sales displays and an increased likelihood of smoking, they consider that the evidence on the impact of Display Bans remains limited. Hence, the introduction of a Display Ban in England, starting with large shops in April 2012 and extending to the small retailers “*which occur typically in the locality of schools*” in April 2015 (p. 694), supposedly provides them with an opportunity for supplementing the existing studies.
88. In particular, the authors “*prospectively*” investigate the association between exposure to POS displays, the development of susceptibility to smoking, and the uptake of smoking “*in the period leading up to the first stage of PoS prohibition in April 2012*” (p. 694).
89. The authors do not advance any theoretical framework or formal hypotheses, and their review of the literature is relatively brief.

Appropriateness of research design

90. The research design employs two cross-sectional surveys of secondary schools in Nottingham, the first in March 2011 involving 11 schools and the second in March 2012 (immediately before the Display Ban) involving 8 of the original 11 schools.¹⁹
91. As the authors note later in their paper, this is a clustered sample in that students are surveyed within their class and school and therefore their responses are not independent of each other. In these situations, the most appropriate design would incorporate the class and school clusters into a multi-level design and/or take measures of the classes and schools as well as the students. This then allows the researcher to identify the effects of interest to them after controlling for class and school influences and the resulting lack of independence in the sampled data. Unfortunately, the authors note that they did not have adequate samples to use multi-level approaches and were forced to employ a post-hoc estimation technique to correct for this lack of independence. Arguably, it would have been better to consider this at the design stage.

¹⁹ The study by Spanopoulos et al. (2014) reports on the first wave of these surveys.

92. The authors do not discuss any possible confounds to their studies, such as media coverage of the Display Ban (particularly around the time of the second study) or any significant changes to cigarette prices over the 12 month period.

Representativeness of the sample

93. The authors obtained the consent of the head teacher and “opt-out” consent forms were distributed to parents of children aged 11 to 16 years. Overall, the 2012 study achieved a 69 percent response rate, which is potentially a strong point of this study, but unfortunately there is no discussion or analysis in the paper of how this varies by class or school. I would suggest that this is an important qualification on their results in that if some classes and/or schools have high response rates, and others lower ones, this: (a) complicates any adjustment for lack of independence; and (b) might raise issues of non-response bias for some classes or schools. The other study using data from the first wave of these surveys (Spanopoulos et al. 2014) reports that only 11 of 36 schools in Nottingham agreed to participate in the study, which reinforces my point here (see paragraphs 493 to 497 for further discussion).
94. The authors then link the completed 2012 questionnaires back with those from 2011, obtaining (after the exclusion of some with missing or inconsistent data) 2,270 students who completed both. Hence these data form a two-wave panel study, a point the authors do not discuss in any detail but which is also a potential strength of their work. This is because measures of the same person at two points in time allow more robust estimates of change than measures of different samples of people at each time, since we do not have to take account of the differences between samples. Another way of putting this point is to say that the response of each respondent at time 1 becomes a baseline control for their response at time 2.
95. There are of course other methodological issues that researchers using panel designs need to consider, such as any biasing effects of repeated measurements and also whether the panel respondents are different to those who only answered at one point in time. These issues are also not discussed in the study, which is unfortunate, as with the data they have and a few simple analyses, the authors could have demonstrated the strength of their panel data. Without these analyses I can only speculate that any potential bias from repeated measurement is probably not an issue (given the 12 months between surveys) but the difference between panelists and others might be. I note the authors had 3,672 completed surveys in 2012 but were only able to match 2,270 to 2011 respondents. A simple comparison between the 1,402 who only completed the 2012 survey and the 2,270 who completed both surveys could have addressed this potential limitation.
96. That said, having two measures of 2,270 students 12 months apart provides a solid platform for their subsequent analyses.

Measurement issues

97. The survey measures relating to smoking are self-reports of smoking among family and friends, smoking status and smoking susceptibility. The authors also included measures of self-perceived academic performance,

rebelliousness, gender and age and, via the student's postcode, were also able to associate an area-level measure of socioeconomic status to each student.

98. The measures of exposure to POS displays were also self-reports, of how often students go to supermarkets and small shops, how often they noticed tobacco displays and which popular brands they recognize. I note the authors do not provide any detail on how "tobacco displays" were defined to the respondents, so the accuracy with which effects are ascribed to actual POS displays is hard to evaluate. Furthermore, there is a potential for these questions to be leading ones. The actual question is not included in the paper but the text indicates it was a direct question focusing on cigarette displays rather than indirectly about what the respondent noticed in the retail outlet. Carter et al. (2015) have pointed out the potential biases that might arise from such direct and leading questions.
99. To "*avoid categories with small numbers*" (p. 694), the authors manipulate the data from several of these measures, either forming binary variables from a split on the median response or otherwise grouping categories together. This suggests a lack of adequate pretesting of the questionnaire measures. It also means some information on variability has been lost by the grouping process.
100. The two measures of possible interest to my review are those of smoking status, a self-reported behavior, and self-reported smoking susceptibility, a potential precursor to behavior but not a behavior in and of itself.
101. For *smoking status* the authors used a question derived from another national survey which classifies respondents into "never smokers" and "ever smokers", where the latter category aggregates those who have tried smoking in the past with those who smoke currently. In my opinion, this is an unhelpful aggregation, since it confuses two quite different behavioral states and therefore makes it harder to establish associations with exposure to POS displays, which could potentially have different impacts on people in the two different states.
102. The questions on *smoking susceptibility* are from a previously validated scale²⁰ built from three questions leading to a binary classification into susceptible or non-susceptible. While the use of three questions to determine susceptibility potentially provides a strong measure, there is clearly an issue of socially desirable responses, since all three questions focus on cigarette smoking.
103. Overall, measurement is a significant weakness with this study, as there are the usual limitations of self-report, coupled with lack of clarity on definitions and questions that are potentially leading or could provoke socially desirable responses. The aggregation of distinctly different behaviors into one smoking status measure is also a weakness of the study—it would be far better to separate current smokers from those who had tried smoking in the past,

²⁰ Pierce J.P., Choi W.S., Gilpin E.A., Farkas A.J., Merritt R.K. Validation of susceptibility as a predictor of which adolescents take up smoking in the United States. *Health Psychol* 1996; 15: 355–61.

particularly as, given the focus of the study, the authors are mostly interested in impacts on current smokers.

Appropriateness of analysis and key study results

104. The authors make good use of their panel design, forming four outcome variables from their measures: (i) susceptibility to smoking in 2012 amongst those who were “non-susceptible never smokers” in 2011, (ii) “ever smoking” in 2012 among those who were “non-susceptible never smokers” in 2011, (iii) non-susceptibility to smoking in 2012 among those who were “susceptible never smokers” in 2011 and (iv) “ever smoking” in 2012 among those who were “susceptible never smokers” in 2011. The strength of the panel design is that it allows such changes in individuals to be clearly identified and associated with the explanatory variables in ways that are not obscured by sample differences, as would be the case in a simple two-survey design with independent samples.
105. The explanatory variables the authors use are their measures of shop visit frequency, noticing POS displays and number of brands recognized, together with what they call “suspected confounders.” For the latter they use a number of formulations of friend and parental smoking behaviors, which potentially might also associate with their outcome variables, as well as age and gender, perceived academic performance and rebelliousness. Again they use measures made in 2011 to explain outcomes in 2012.
106. The authors also combine some of their measures to “estimate joint effects” (p. 695)—this step is not particularly explained or justified in the paper and, perhaps inappropriately, combines measures of different types into one. For example, frequency of noticing POS displays is combined with number of brands recognized. What results is an overall categorical variable with the various combinations of the levels of each constituent variable becoming a new category e.g. “Notice displays sometimes or less and recognize 1 to 5 brands.” This approach has the benefit of being easier to interpret than the more common approach of adding interaction terms to the regression models. On the other hand, there is no justification or analysis in the paper supporting the appropriateness of combining variables in this manner.
107. Another limitation to their analyses is that they include respondents with missing values on the explanatory variables in their analyses by “*coding missing values as a separate category, to maximize study power*” (p. 695). A better procedure might have been to impute these missing values, however the authors do report the number and impact of these missing values, which is good practice.
108. Finally, given the binary nature of their outcome measures the authors use multinomial logistic regression for their analyses, conducting a number of sensitivity analyses on these and using “relative risk ratios” as their statistic of interest. They also use significance levels of 0.01 in their statistical tests, which is perhaps not strong enough a test given their large sample and my previous discussion of this point. However, I note several of their results are significant at a higher level, notably 0.001.

109. The use of relative risk ratios rather than the odds ratios that are more common in this field needs brief mention. Technically a relative risk ratio adjusts the odds ratio for the level of risk observed in a group who are not exposed to the “risk.” Here the authors define the “not exposed” group as those who were not exposed to POS displays (presumably as they did not visit shops) or did not notice these POS displays. While the use of the relative risk ratio might be preferable to the odds ratio (as it is the more conservative approach) the authors’ categorization schemes (which do not identify the non-exposed groups) and the lack of detail on the smoking behavior of these non-exposed children makes it hard to calculate effect sizes. I will return to this point subsequently. Some of their analyses also do not follow the strict definition of relative risk as they use “infrequently” rather than none as the baseline.
110. **Key study results.** In terms of raw data, the 2011 survey shows 1,576 non-susceptible never smokers, 494 susceptible never smokers, and 200 ever smokers. Of the 1,576 non-susceptible never smokers, 20 percent became susceptible never smokers in 2012 and 7 percent became smokers. Of the 494 susceptible never smokers, 45 percent did not change status, while 26 percent became non-susceptible never smokers and 29 percent became smokers. The authors do not tell us what happened to the 200 students who smoked in 2011.
111. For a potential explanation of these changes, I focus on the regressions which the authors “adjusted” (controlled) for confounding factors, particularly as these confounding factors appear to have effects which are as strong or stronger than those of shop visits and POS display exposures (see Table 1 of the paper on the unadjusted relative risk ratios for the various factors).
112. The adjusted logistic regressions show that those non-susceptible never smokers in 2011 who visit shops, and notice POS displays more frequently and recognize more brands are more likely to become susceptible to smoking by 2012. All these regression coefficients are significant at the $p < 0.002$ or < 0.001 level. However, once adjusted for control variables there was no observed effect on the self-reported behavior of smoking status. For those who were susceptible never smokers in 2011, no significant transition to non-susceptible was observed, and as before there was no observed effect on smoking status in this group either.
113. When the combined variables are used, similar effects on susceptibility were seen on non-susceptible never smokers. Moreover, those who noticed POS displays most or every time and recognized at least one brand were three times more likely to become susceptible than those least exposed. And those who noticed displays most or every time and recognized more than five brands were more likely to have become smokers in the 12 months (relative risk ratio of 3.4, $p < 0.002$).
114. However, these latter results become non-significant in the sensitivity analyses, where the authors drop the parental, friend and family smoking controls from their logistic regression models. In contrast, those susceptible never smokers who recognized more than five brands became significantly more likely to have taken up smoking by 2012 when these controls are dropped (relative risk ratio of 2.1, $p < 0.001$). The authors do not discuss the

implications of these results in any detail, though they argue tobacco marketing may also influence these control variables. That may be the case but, in my opinion, (i) many other factors also influence parental, friend and family smoking and (ii) I believe it is better to keep these controls in the equations so that the effects of shop visits or noticing displays over the 12-month period may be more accurately identified. For these reasons I do not believe these sensitivity analyses are informative.

115. The authors conclude that visiting shops and noticing POS displays is “*associated with an increased likelihood of non-susceptible never smokers becoming susceptible to smoking, but is not related to the likelihood of becoming an ever smoker*” (p. 700, my emphasis added). They also note the combined variable results above but point out, correctly in my opinion, that for these they are not able to determine whether these effects are due to the POS display or to the brands themselves.
116. The other limitation they note is that the actual number of children changing their levels of susceptibility or smoking status was small and so they conclude their study has low power. This comment refers to the fact they were unable to contrast the effects of large versus small retailers, or to estimate the independent effects of shop visits, POS displays and brand recognition more robustly. Finally, they also note their measures are self-report and “*hence relatively open to error and bias*” (p. 700).
117. **Effect size.** As noted above, the lack of detail in the paper make it difficult to convert the relative risk ratios to the more common odds ratio and therefore to make a definitive assessment of effect size. However, if we can assume the risk in the non-exposed group is low then the risk ratio can be approximated by an odds ratio of similar magnitude.²¹ In which case the risk ratios reported above, e.g. 3.4, represent effects somewhere between the medium and large magnitudes recognized in the literature.

Overall conclusions

118. The use of measures taken in 2011 to identify and explain changes in the smoking susceptibility and smoking status of individuals over the 12 months to March 2012 is a strong point of this study, as is the relatively large sample of school students and the high response rate for the surveys. Large sample panel designs such as this provide stronger evidence on causality, particularly here where the authors use 2011 responses to control and predict 2012 outcomes.
119. However, the self-report measures and the possible errors and biases from leading questions, and/or questions that might encourage socially desirable responses, place limitations on the robustness of the results obtained. Similarly the lack of any analyses comparing those who provided data on two occasions

²¹ Viera AJ (July 2008). “Odds ratios and risk ratios: what’s the difference and why does it matter?”. *South. Med. J.* **101** (7): 730–4

versus those who only provided it once might also place limitations on the generalizability of these results.

120. That said, the authors find no significant association between the frequency of shop visits, or of noticing POS displays, and smoking behavior. They do find a significant association of these factors with smoking susceptibility, with a medium to large effect size. Effects on behavior do appear when the combined variables are used but as the authors themselves note it is difficult to determine whether this is due to the POS displays or brand recognition. Overall, in my opinion, this suggests the correlation between the proposed precursor to behavior, smoking susceptibility, and smoking behavior is, at best, small in magnitude, limited to those with high brand recognition and driven by many factors other than the POS displays themselves.
121. However, from the perspective of the objectives of my report, the major limitation of this study is that it occurs before the Display Ban is implemented. Thus it cannot provide any evidence on what the impact of a Display Ban will be on either susceptibility or behavior, since the critical event has not yet happened. The authors argue that removal of POS displays “*might play an important role in reducing smoking uptake and prevalence among young people by reducing the numbers who become susceptible to smoking*” (p. 701). However, this is an indirect argument and, given the limitations of their study and the lack of any significant association between shop visits, exposure to POS displays and smoking behavior, one that, in my opinion, is difficult to sustain.

Suzan Burton, Lindie Clark, and Kristina Jackson. The association between seeing retail displays of tobacco and tobacco smoking and purchase: findings from a diary-style survey. *Addiction* 107.1 (2012): 169-175. (Burton et al., 2012).

Objectives, background and theory

122. The authors state their aims are to “*assess the impact of retail displays*” on smokers purchasing behavior, including cessation attempts (p. 169).
123. The authors review the context and prior research in the field, highlighting a number of theoretical and methodological issues that might place limitations on the conclusions that can be drawn from this research. First, they discuss the possibility of “self-attribution” bias in some studies (“*the tendency of individuals to assume less personal responsibility for failure than success*”, p. 169). This bias might mean POS displays are reported as influencing purchases when in fact these purchases would have occurred anyway. Second, they suggest that those individuals who visit stores more frequently (and are thus potentially more exposed to displays) may differ in systematic ways from those who visit less. Hence these (unobserved) differences may create spurious correlations between exposure and smoking behavior. Third, they note issues of causality and the precise mechanism by which retail displays work. For example, displays might simply remind smokers to make a purchase and not necessarily lead to higher consumption—a finding which would be consistent with marketing research on the effects of promotions in general, many of which simply shift purchases in time rather than increase consumption *per se* (see paragraph 51 earlier).
124. The authors’ overall conclusion on the research they cite is that this shows only that smokers *believe* retail displays lead to impulse purchasing but does not actually establish that such displays lead to higher levels of purchase or smoking (p. 170). Their goal is thus to use a diary-style methodology to examine exposure to retail displays, purchasing behavior and consumption over time.
125. The authors provide no theory or hypotheses as such, moving directly to outline their methodology after their discussion of the limitations of prior research.

Appropriateness of research design

126. It is important to note that this study was conducted in NSW Australia with data collected from October 2007 to March 2008, which is just before retail displays of tobacco products in NSW were subject to certain restrictions. Primarily these restrictions limited tobacco products to one display at the POS. However, the fact that the study was done in the period before these restrictions means that measurement of POS displays needs to separate their effects from any other tobacco displays or materials in the store environment. The authors apparently do not do this so in my review here I use their terminology of “retail displays” rather than POS displays.

127. The time period of the study also means that there may have been some media coverage on the changed regulatory environment, potentially influencing smoking behavior and potentially confounding study results. However, as this was not a Display Ban as such, this coverage may have been relatively limited.
128. The instructions for the diary asked respondents “*to record a number of measures for each 4-hour interval they were awake over a 4-day period*” (p. 170). Different respondents were assigned different start days so that all seven days of the week were covered by the study.
129. While the 4-day focus of this design imposes a reasonably light burden on the respondent, it does not really allow a substantive assessment of consumption, especially when compared with the behavior of a smoker over longer-time periods or the time it might take someone to quit smoking. Nor does it allow a strong assessment of causality, since the 4-hour time periods would presumably include both exposure to retail displays and any cigarette purchase within the same time period (i.e. store visits are typically shorter than 4 hours). Indeed, the authors note this point themselves, suggesting that future research use 1-hour recoding periods. In my opinion, even this would not be short enough, as others in this field have reported that the typical visit to a convenience store is more in the order of 15 minutes (Kim et al., 2013).
130. The authors do not discuss any potential panel bias in their study, bias that may arise from respondents focusing and reporting on their smoking behavior for four days. Arguably better designs might have been a diary with tobacco products as only one area of exposure and behavior for respondents to report on out of several product categories and/or the inclusion of a control group with different reporting requirements.
131. Overall, I am led to the conclusion that while a different and interesting approach to the typical survey in this field, this study, in reality, does not have a better design for “assessing the impact” of retail displays than the typical survey. As far as causality is concerned, at best it can only show an association between exposure to retail displays and smoking behavior.
132. Where this design does offer some potential is that it potentially provides a description of behavior over more purchases and behaviors than those studies that look at only one purchase, and it gives us a more detailed picture than those studies that ask simply for recall of the respondent’s aggregate behavior in the past.

Representativeness of the sample

133. As the authors note, their sample is a convenience one and, since their focus is on smokers, for whom no sampling list exists, they had to sample more broadly and then exclude non-smokers. They rightly conclude it is difficult to generalize from this sample to the population of smokers (p. 174).
134. From an initial list of 20,000 adults, 3,735 individuals responded of whom the usable sample of smokers was 1,109. While the authors claim this number is close to what you would expect given smoking levels in NSW as noted there is no guarantee that these 1,109 individuals are representative of the smoking

population. The authors do look at early versus late responses (a common test of non-response bias) and find no significant differences, but I would also have liked to see more detailed follow-up research on reasons for non-response to assess the criticism the authors themselves made of other studies, namely selection bias.

135. Overall, given we cannot generalize from this sample; it is difficult to make policy recommendations from the results of this study.

Measurement issues

136. For each 4-hour period, respondents were asked to record their physical location, the presence of family, friends or others smoking, whether they saw cigarettes for sale, any purchases of cigarettes and number of cigarettes smoked. Various demographics were also collected.
137. Exposure was measured by the simple question “During the past 4 hours did you see any cigarettes for sale anywhere?” Given that in 2007 and 2008 cigarettes displays were not yet restricted solely to the POS this question covers any and all display of cigarettes in the store, which is presumably why the authors use the terminology “retail displays” rather than POS displays.
138. Whether these measures can overcome attribution or social desirability biases is open to question, since the authors include no measures or analyses to address these biases and, in the end, the measures obtained are self-reports and therefore subject to such biases, along with inaccurate recall.
139. Overall, the authors’ lack of attention to the validity and reliability of their measures, and to any potential biases in them, is a concern, particularly as the authors partially motivate their study as overcoming some of these limitations.

Appropriateness of analysis and key study results

140. The authors use logistic regressions for their yes/no measures (smoking and purchase) and count data models for numbers of cigarettes consumed. They correctly use multi-level models as they have repeated measures from each responding individual. Family, friends and others smoking and various demographics are used as control variables in these models.
141. Despite it being part of their stated aims, the authors do not model those attempting to quit smoking separately, rather they include quitter versus smoker as one of their control variables. This choice is not explained in any detail and results in a coefficient in their two main models that is much larger than any of the other coefficients and possibly highly correlated with the dependent outcomes for definitional rather than substantive reasons. That is, there is a risk that essentially the same measure has been introduced on both sides of these equations. This could particularly be the case for the smoking model. I would have liked to see analyses presented without these smoker/quitter controls to assess whether this is the case or not and also to establish whether the other effects are robust to including/excluding this control.

142. **Key study results.** In terms of raw results some 998 respondents were smokers and some 111 attempting to quit smoking. A “large majority” (88 percent) reported smoking during the 4-day period. As would logically be expected, exposure to retail displays varied according to whether respondents were at home or not. Hence the authors separately examine those periods when the respondent was outside the home, finding that they saw cigarettes for sale in 43 percent of the 4-hour “out-of-home” periods.
143. Before presenting their regression results the authors discuss the problems of causality in their data (partially caused by the 4-hour reporting periods as noted above but also due to the nature of the exposure question asked). As they correctly note, it is impossible to say whether a purchase in a 4-hour period is a result of seeing a display in that period or a consequence of buying cigarettes, which necessarily involves seeing them. Furthermore, they state that people who buy more cigarettes are more likely to consume more than people who buy less: “[T]hus seeing cigarettes for sale could be an outcome of deciding to buy cigarettes and any association between seeing cigarettes for sale and smoking may be caused by heavier smokers being more likely to buy cigarettes” (p. 172). The authors therefore conduct their modeling exercises only on time periods when the respondent saw cigarettes for sale but did not buy them. They also conduct analyses with lagged variables where they associate tobacco purchase in one 4-hour period with seeing cigarettes for sale in the previous 4-hour period (controlling for any purchases in the previous period).
144. The logistic regression for smoking and seeing cigarettes for sale for those time periods where the respondent did not buy cigarettes showed a significant association between the two ($p < 0.001$), after controlling for other factors. Similarly, the analogous regression between the number of cigarettes smoked and seeing cigarettes for sale was also significant ($p < 0.001$). The various control factors were also significant at the same level, except for socioeconomic status in both regressions, and gender in the smoking regression. The lagged models are not significant ($p = 0.066$).
145. It should be noted that the unit of analysis here is smoking or consumption events and the sample yields over 20,000 events from 1,082 respondents. Hence it is also not surprising that many effects are significant and the more interesting questions concern the size of these effects.
146. **Effect size.** In the smoking regression the odds ratio for seeing cigarettes on sale is 1.5 and for the number of cigarettes smoked it is 1.2, which are both positive effects of small magnitude. (The latter ratio of 1.2 is strictly an event rate ratio rather than an odds ratio but I believe my interpretation to be valid). As the authors note these magnitudes are somewhat similar to those for seeing friends, family or others smoking.

Overall conclusions

147. The authors argue that excluding 4-hour periods where cigarettes were both seen and bought by respondents provides a “*very conservative analysis*” since it effectively excludes any increase in smoking due to impulse purchases (p. 173). Here they imply the real effect of seeing cigarettes for sale on

consumption is much larger. In my opinion, this argument is an over-stretch. It would be more accurate to say that their analyses say nothing about impulse purchasing; rather they show an association between seeing cigarettes and smoking and consumption of small magnitude. However, this effect is also confined within the 4-hour period, so while I commend the authors attempt to identify causality, it might also be possible to argue that this remains unproven. For example, did the respondent increase their consumption because they saw cigarettes for sale or did they see cigarettes for sale because they just consumed one? Without a time sequence of events within the 4-hours we cannot know.

148. I also have reservations about the reliability, validity and potential biases of the measures in this study that, together with its limited ability to generalize to the population of smokers because of the convenience sample, make it difficult for me to believe it provides strong evidence in favor of a Display Ban, which is what the authors suggest in their last sentence. In addition the study does not consider minors and also does not specifically and directly measure either initiation or cessation as behaviors, only purchase and consumption. It therefore does not provide evidence relating to the objectives of my report.

Owen B.J. Carter, Tina Phan and Brennen W Mills. Impact of point-of-sale tobacco display ban on smokers' spontaneous purchases: comparisons of post-purchase interviews before and after the ban in Western Australia. *Tobacco Control*, 24.e1 (2015): e81-6. (Carter et al., 2015).

Objectives, background and theory

149. The authors state that their objective in undertaking this study is to assess the impact of a Display Ban in Western Australia on spontaneous purchase behaviors. They do not posit a theoretical model for this impact, but discuss the background to their work, citing previous work (mostly on young adult smokers). They then go into methodological criticisms of their previous work in this area and that of others (in particular, issues around recall and response bias and the need for control groups). Thus, the focus of this study is primarily improved methodology from their previous study (Carter et al, 2009).
150. The authors (and others) have previously used exit interviews to minimize the problems of recall bias often seen in surveys. That is, asking people what they did over weeks or months before an interview can lead to inaccurate recall, as people do not necessarily memorize the details of minor events in their life. Instead, by asking questions immediately on exit from a store, researchers can potentially reduce such biases to a minimum.
151. However, exit interviews as such only address one potential bias, and the authors note their previous work using this technique (Carter et al, 2009), has been criticized by others for using "leading" questions (Keegan, 2009). In the paper, they present the problematic question, "*Did the cigarette pack encourage you in any way to purchase cigarettes in this instance?*". This was asked immediately after smokers had stated they made a spontaneous purchase. In my view also, the question is clearly leading respondents to "explain" their purchase as due to the "encouragement" of the pack.
152. The paper then discusses a more recent study (Clattenberg et al, 2013), which uses indirect questions, such as "*Please state which of the following influenced your decision to purchase cigarettes in this instance?*", to try to get round the problem of leading the respondent to a certain answer. Clearly, this is a better methodology since the respondent can reach their own conclusion by selecting from a list of possible factors that may influence them. However, the authors note that this study, despite improved methodology, did not achieve statistically significant results for any association between the impact of POS displays and spontaneous purchase.
153. Finally, the authors note their 2009 study (and those of others) was also criticized for not having a control group, which might have allowed them to show a difference between the levels of spontaneous purchase in the absence as well as presence of tobacco displays (Keegan, 2009). They cite a four-country study (Li et al. 2013, which I discuss in detail below) as providing some circumstantial and supportive evidence for their previous findings, in that smokers more often recall making spontaneous purchases influenced by displays in countries without bans than in those in countries with bans. However, they note that the Li et al study does not distinguish between

tobacco displays and in-store advertising of tobacco products, making it impossible to distinguish the impact of the display of tobacco products alone.

154. In passing, I would note that the authors' introduction to this paper indicates that the recalled levels of "spontaneous" purchase in previous studies are relatively low (approximately 10 percent of smokers), implying that any researcher who wishes to study this smoking behavior needs either a large sample or extensive screening techniques to obtain adequate subjects for study.
155. The authors use this background as their motivation for their current study, which uses the introduction of a display ban in Western Australia in 2010 as an opportunity to implement improved methodology. In particular, they employ a quasi-experimental design comparing post-purchase exit interviews before and after the ban and using indirect questions to attempt to overcome the recall and response biases in their previous work. The before and after ban format also provides them a quasi-control group in that they obtain respondents in both the presence and absence of displays.
156. In conclusion, the authors posit no theory or hypotheses as such; their aim is to deploy a more advanced methodology than was the case in their 2009 study.

Appropriateness of research design

157. The authors utilize a before-after design: samples were taken both two months before and two months after the Display Ban in Western Australia took effect. They chose to sample two months before the ban to avoid any confounding that might result as retailers began the changeover in their shops a few weeks ahead of the Display Ban. However, the authors do not discuss other possible confounding factors, such as media or information campaigns related to the introduction of the Display Ban. Quasi-experimental field experiments such as these are less controlled than laboratory studies and therefore vulnerable to other events in the environment. In this case, particularly given the short time frame after the ban, any observed change in behavior post ban might be due to the absence of displays in stores or the surrounding media environment? Equally, in my opinion, it is hard to know to what extent the pre-ban benchmark is a true one or is altered in some unknown way by media commentary?
158. In addition to any potential confounding factors in the environment, there are potential issues of generalizability around the recruiting process for those interviewed. The nature of the design is that only those observed purchasing cigarettes on the day are approached for an interview. While this minimizes recall bias, and is a strong point of the study, it makes it difficult to generalize to the population of smokers at large because we do not know how representative the sample is of smokers in general. To be fair, the authors acknowledge this point as a limitation of their work, noting that their results cannot be extrapolated to cigarette consumption as a consequence.
159. Finally, the authors themselves point out their samples are convenience samples, not random ones, nor are they randomized across the pre/post ban

conditions as would be the case with a true experiment. Hence, variations between the pre and post-ban results may also be due to sampling biases.

Representativeness of the sample

160. The sample for this study is made up of adult daily smokers, with a total sample size of 402 (220 interviewed before the ban and a different 182 after the ban). The authors select supermarkets in four suburbs of Perth, covering different socioeconomic status, though with some bias to more affluent areas (two of four supermarkets). The pre and post-ban samples have similar profiles on gender, age and consumption.
161. Participants were recruited on exit of the store by observation of a purchase. It is worth noting that the authors do not mention whether any of the participants knew or noticed they were being observed, which might lead them to “prepare” their answers. They were asked to participate in a five-minute survey on “health issues”, and 68 percent completed this survey (which is a good response rate, lessening concerns about non-response bias). Responses were also screened so that the sample was made up of adults over 19 who smoked at least one cigarette per day. However, one potential concern is that while the cover story was “health issues” the first questions on smoking patterns would immediately signal to the participant what the survey was about, and increased the possibility of socially “desirable” responses.

Measurement issues

162. “Spontaneous” purchase is defined in this study as whether the respondent intended to buy before entering store or not, asked by a simple question on exit. The authors then asked free and cued recall questions on factors influencing their purchase, the free question being “What did you see, if anything, in the store that prompted you to purchase cigarettes in this instance?” and the cued recall question “Please state which of the following influenced your decision to purchase cigarettes in this instance” (list of five items, rotated to avoid order effects). For spontaneous purchasers these questions are a conscientious attempt to avoid leading the respondent. However, for intentional purchasers (those who intended to buy cigarettes before they entered the store) they rather focus attention on store factors as opposed to non-store factors.
163. While acknowledging the need to keep the survey time short (five minutes), it would possibly have been useful to approach the cigarette questions with an initial question on motivation for shopping that day. It is also my view that it might have been useful to include a concluding question on whether the respondent recognized the intent of the survey (as a possibly indicator of social desirability bias in their answers).
164. The authors note the fact that free recall of displays is low in their surveys (5 percent pre-ban) whereas cued recall is higher (24 percent pre-ban), which is a potential limitation to their conclusions. They argue their cued results are consistent with other studies but given the immediate nature of the exit survey it is arguable that the free recall results show displays are not top of mind for most respondents and therefore less likely to be a factor in their purchase.

Appropriateness of analysis and key study results

165. The authors' employ Fisher exact tests for pre- versus post-ban measures, which is appropriate given the small cell sizes they have for several analyses.²² They also use these tests for comparisons of spontaneous versus intended purchases and younger versus older adults. They use one-tailed tests because of their (informal) hypothesis that the impact of displays would be lower after ban. And they present confidence intervals on their charts, which is good practice.
166. **Key study results.** After the ban, fewer smokers reported making spontaneous purchases (20 percent versus 28 percent, $p < 0.05$) and fewer claimed displays influenced their purchase (free recall 1 percent versus 5 percent, $p < 0.05$; cued recall 4 percent versus 22 percent, $p < 0.001$). After the ban, fewer *spontaneous purchasers* nominated the influence of displays (free recall 6 percent versus 10 percent, not significant, cued recall 11 percent versus 40 percent, $p < 0.01$).
167. The authors mostly cite 0.05 probability levels for their one-tailed tests, which is a relatively weak level of significance, though some results are stronger, particularly those involving cued recall or planned purchasers. The authors provide one post-hoc power analysis in an, in my opinion unsuccessful, attempt to imply the 50 percent reduction in spontaneous purchasers mentioning displays in free recall pre/post ban is a useful finding. However, this result came nowhere near attaining any meaningful level of significance and the power analyses demonstrates the sample had little statistical power to detect such a change.
168. **Effect size/behavior.** The key result on *behavior* is for self-reported spontaneous purchase, which fell from 28 percent before to 20 percent post the introduction of the Display Ban, a fall which is significant at the $p < 0.05$ level. However, following paragraph 41 above, the odds ratio of 20/28 or 0.7 denotes an effect of *small* magnitude.
169. In my opinion, the real limitation of the study is that the sample includes only 62 spontaneous purchasers before the introduction of the Display Ban and only 36 after. Given the low frequency of mentions of displays, any apparent lowering of the impact of displays thus rests on small numbers of respondents. Before the ban, free recall of displays by spontaneous purchasers is based on the answers of four people, cued recall on 25 respondents. These are small subsamples from which to build substantive conclusions. Other analyses might have similar problems of small subsample sizes (see the small percentages in many results). It is true the results for planned purchasers are more robust but they were not the main objective of the study and there are also issues of whether they were overly primed to focus on displays (see above).

²² The Fisher Exact test is a statistical test used in the analysis of contingency tables. It is often used when small numbers of observations in each cell of the table make the use of the more common chi-square test inappropriate.

Overall conclusions

170. The authors show the way forward by overcoming several problems of recall and response bias, with commendable attention to the formulation of their questions. Unfortunately, the convenience nature of the two samples makes it difficult to extrapolate their findings to the universe of smokers, and the non-random nature of these samples also makes it harder to make valid pre/post ban comparisons. Critically, the low statistical power of their analyses also casts some doubt on the validity of several of their results and conclusions.
171. This study does not address minors or measure initiation, since those interviewed were adult smokers 18 years or more of age. In respect of smoking cessation in minors and adults, the reported reduction in spontaneous purchasing reported in exit interviews and correlated with less perceived influence of displays might potentially provide circumstantial evidence on this objective. However, smoking cessation was not specifically measured, we do not have a full picture of the consumer's purchasing behavior, only one event, and the study has small sample sizes. These limitations do not allow me to draw any conclusions on whether a Display Ban is likely to achieve the objectives outlined in paragraph 6(a) and (b) above.

Sally Dunlop, James Kite, Anne C. Grunseit, Chris Rissel, Donna A. Perez, Jane Young and David Currow. Out of Sight and Out of Mind? Evaluating the impact of point-of-sale tobacco display bans on smoking-related beliefs and behaviours in a sample of Australian adolescents and young adults. *Nicotine & Tobacco Research* 17.7 (2015): 761-768. (Dunlop et al., 2015)

Objectives, background and theory

172. The stated objective of this study is to look at medium term impact on youth smoking of the Display Bans introduced in NSW and Queensland, Australia, in July 2010 and November 2011 respectively. The authors source their data from an existing yearly survey on the impact of tobacco promotion conducted in 2010, 2011 and 2012.
173. The authors suggest that evidence in the literature shows that the display of tobacco products can communicate brand imagery to children, and this is associated with increased susceptibility to smoking and smoking among youth. They also state that published experiments show youth exposed to images are less supportive of control policies, more likely to recall brands, and that the display of tobacco products at POS stimulates impulse purchases and encourages cigarette consumption.
174. The authors provide no background theory or model of the impact of tobacco displays, but do posit two hypotheses based on the prior research they have selected and reviewed, namely that: (1) those interviewed after the Display Ban would have less risky smoking-related cognitions and behaviors than those interviewed before; and (2) the difference in outcomes would be most apparent in those who visit stores selling tobacco most frequently.
175. The authors do not advance any hypotheses as to how cognitions and behaviors might evolve over time, which is a potential oversight given they have three waves of data and the bans themselves were introduced in a staged manner. However, the study demonstrates that unfortunately the authors do not have a clean set of “before” Display Ban data for NSW and hence it is difficult to extract clear trends (see below).

Appropriateness of research design

176. The authors’ Wave 1 data is from June 2010 and they note at that time neither NSW nor Queensland had “fully implemented” a Display Ban. However, NSW actually banned displays in large retailers from January 2010, and in small retailers in July 2010. By the Wave 2 survey, in June 2011, a Display Ban had been introduced in NSW in most retailers for some 12 months, but was not yet in effect in Queensland. By the Wave 3 survey in June, 2012, the Display Ban in NSW had been in effect for 24 months and for 7 months in Queensland.
177. In my opinion, the Wave 1 for NSW is not strictly before the introduction of the Display Ban, as there must have been press and surrounding discussion at the time of the survey (given the imminent ban in small retailers), and the display of tobacco products was already banned in large retailers, which are presumably an important source of supply of tobacco products (see Quinn et

al, 2011) and are visited by many consumers on a weekly basis. This does not necessarily invalidate the results of this study, but requires separation of the NSW analyses from Queensland, which the authors do not do. Instead, they suggest in the conclusions that this time overlap may have attenuated changes after the introduction of the Display Ban in NSW. However, it is possible that some other factors might have been at work and/or that there is no valid “before” ban measurement in NSW.

178. In the conclusions section, the authors note that their study design would be better with a control group. This observation is correct and noted by others in the literature (e.g. Carter et al., 2015). This is because of a classic problem in “quasi” field experiments that many unobserved factors are at work in the environment and it is difficult to identify the impact of the specific factors of interest to the researcher. The utilization of a control group for comparisons gives greater confidence that any different effects observed in the experimental group are those hypothesized and not confounded with these unobserved factors. The authors argue their specific analyses (use of covariates, split samples, etc.) go some way to overcoming this problem, a point I will examine later.

Representativeness of sample

179. The surveys themselves were conducted by random digit dialing telephone methods, and participants within household also randomly selected. Permission was sought for those aged 15 and younger. While this is a clear attempt to get the random samples necessary for experiments, there are methodological questions in this instance, namely their use of landline directories as the sampling frame. Two issues potentially emerge. First, many households are now mobile only and may therefore be omitted from the sample. Second, busy people are often not at home or less prepared to answer their landline phones. Indeed, the authors note that 33 percent of the target 12-24 age group lives in mobile only households. In defense, the authors argue their samples are representative on other factors and, if it exists, this bias “was consistent across years”. This is a plausible argument given the short time period between each survey wave.
180. The response rates at each wave were 45 percent, 42 percent and 39 percent respectively, and the authors did not indicate whether this differs between NSW and Queensland. These are adequate response rates though it would have been desirable to examine the reasons for non-response in more detail to identify any other possible sources of bias. The authors themselves assess their sample as “relatively consistent” with population data, but note some age-related discrepancies that they addressed with “post-stratification” weights. This is a legitimate procedure that allows them to claim more generalizability to the population at large. Weighting can reduce the power of statistical analyses (among other problems) but this is probably not of concern here as the authors have large samples and their corrections are small.
181. Across the three waves, the sample size was large at N=6,014, although the numbers drop across the three waves (N=3007 Wave 1, N=2007 Wave 2 and N=1000 Wave 3). That said, the authors have good statistical power for their

analyses and their minimum cell size is 87 in Wave 3 (Table 1) and much larger for the other waves.

Measurement issues

182. The authors outline their measures in some detail. As they do in their paper, I separate this discussion into outcome or dependent variables, independent predictors and covariates.

- (a) **Dependent variables.** Respondents were asked to provide their own recall of seeing tobacco displays in the last month. As noted by Carter et al., (2015), this is unlikely to be a useful measure due to recall biases. Brand awareness was measured by asking for names of brands, with most only being able to name 1 or 2 so the authors formed this into a binary variable (none or one or more). Normative beliefs about the behavior of peers were measured by asking respondents to estimate how many “people their age” out of every 10 they thought smoked. The authors then compared this to national age-group smoking statistics to compute an over/underestimate of smoking as a dichotomous variable. This is a questionable variable transformation because it is driven by averages whereas there is presumably heterogeneity in the population. Thus any one individual’s over/under-estimation against the national average might not be an over/under-estimation against their peer group. It is not clear why the authors did this and I would also have liked to see the original variable analyzed alongside the transformed one.

Finally, the survey asked how often seeing tobacco displays made them think about smoking and, for current smokers, how often they purchased cigarettes on impulse due to seeing displays. Following Carter et al., (2015), these are arguably leading questions as the respondent is pointed to displays as the “cause” and given no alternatives.

One concern is that the authors dichotomize nearly all of their dependent variables. They do not provide much by way explanation or justification for doing so. Dichotomization discards important information on variation and restricts the analytical methods that can be used on these data. It may well be that this step was forced on them by the distribution of people’s answers to these questions (e.g. some answer categories not used) but the point needed discussion as it might indicate problems with the formulation of these questions.

- (b) **Independent Predictors.** The sole independent and predictive variable is the stage of ban, namely a classification of the data into three levels corresponding to the survey waves: (1) before the introduction of the Display ban (but note the discussion above as to whether this is truly before the ban in NSW), (2) 6-12 months after the introduction of the Display Ban and (3) 24-months after the introduction of the Display Ban. The authors also mention that there is no data for Queensland 24 months after, so the third level is solely based on NSW.

- (c) **Covariates.** The authors include a number of other variables as covariates; namely demographics (age, gender and state), self-reported visit frequency for four types of store, self-reported exposure to anti-smoking advertising, and smoking exposure and experience variables. “Exposure” is computed from the number of current smokers in the household and how many of their five closest friends smoke. “Experience” combines exposure with the respondent’s current smoking behavior.

Appropriateness of analysis and key study results

183. Given that the authors dichotomized most of their dependent variables, they necessarily have to choose an appropriate analytical method for binary data, here logistic regression. The authors run a number of such models—one on the overall data, two models on subsamples formed by frequency of store visitation which they treat as a sensitivity analysis, and a final model to predict self-reported impact of displays, which they say has a smaller sample and for which they do not present full details.
184. The authors appear to run separate models on each of their dependent variables. A multivariate statistician would likely raise the objection that these dependent variables are presumably correlated with each other. Failing to account for these correlations in the analysis may reduce its statistical power, mask important relationships in the data and potentially violate some of the assumptions of logistic regression. Of course the dichotomization of the data may have reduced these correlations but then only at the expense of loss of information.
185. The discussion of the store visitation frequency “sensitivity analysis” is relatively hard to follow. This is an important analysis because presumably visit frequency relates to POS exposure. However, without a theory or model of how such displays work it is difficult to say whether increased visit frequency might lead to greater impact over repeated exposure or alternatively to habituation (“wearout”) and less display impact over time. To empirically determine the effect, one would need adequate variation in store visit frequency. Unfortunately the form of the analysis the authors chose does not lend itself to ready interpretation on these points.
186. **Key study results.** In my view, it is anomalous that “*more than half of this sample of youth*” still reports seeing tobacco displays between 6 and 24 months after the Display Ban was introduced. This suggests one of the measures may not be working as intended. Indeed, other studies I have reviewed in preparing this report demonstrate that retailers were compliant with the legislation of this nature and so we can likely rule poor compliance out as an explanation of this result (e.g. Zacher et al, 2013 found high compliance in Victoria, Australia). The authors suggest that this may be because respondents know cigarettes are still behind the covered display unit and/or the presence of health warnings at the POS. But, they also acknowledge that their measure is potentially flawed. In my view, it also suggests the need to ask more carefully formulated and sharply defined questions around the various attributes of the POS. In summary, while the authors show highly

statistically significant effects of ban stage on POS display recall ($p < 0.001$), given the problems with the measure itself, it is hard to regard this result as fully credible.

187. For their sensitivity analysis, the authors split the sample into two subsamples (medium/low visit frequency) and, in essence, compare the ban stage coefficients from the two models. While the authors show significant effects and argue for the impact of POS displays on both samples, there is in fact little difference between the two groups on the display recall measure. Where there are apparent difference is in medium frequency groups' decrease in overestimation of peer smoking and current smoking behavior. I would argue that this analysis introduces additional confounding issues of subsample differences. The distinction between low and medium visitation is between 0 and 1 or more store visits in the last month, which is hardly adequate to test a varying exposure hypothesis. In fact their use of visit frequency as a covariate in the overall model is probably a sounder analysis, since there are no issues of subsample differences and they use three levels of visit frequency (with the significant difference on display recall being for the medium frequency versus low or high). However, it should be noted that the effect size is small across the three levels, 74 percent versus 68 percent and 69 percent respectively.
188. Finally, Dunlop et al present their results on whether POS displays stimulate thoughts about smoking or lead to impulse purchasing. In addition to doubts about the leading nature of these questions, the results here are either weakly significant ($p < 0.04$) for thought stimulation or not significant for impulse purchasing ($p = 0.21$). Given the sample sizes are respectively 1,750 and 859 and thus there are adequate numbers of respondents recalling displays for both measures (17 percent and 23 percent "pre-ban" respectively), this suggests these effects are not strong or not present.
189. **Effect size/behavior.** The *behaviors* studied here are current smoking and impulse purchasing. The result for impulse purchasing is not significant; hence I focus on current smoking. The effect 7-12 months post ban is not significant; the effect 24 months post ban is, with an odds ratio of .73 (drop from 15 percent before to 11 percent, 24 months post-ban). Following paragraph 41 above, this would be a *small* magnitude effect.

Overall conclusions

190. This study has a solid longitudinal design and large samples sizes such that the effects of a Display Ban should, in principle, be identifiable from these data. However, the lack of a control group, as the authors themselves acknowledge, and doubt as to whether the Wave 1 NSW data is really "pre-ban" place qualifications on their results. Issues around the questions themselves (such as the leading nature of some and the anomalous results on recall) accentuate these qualifications. Further, the manipulation of the original data by the authors into a binary form, without much explanation or confirmatory analyses, along with the failure to take account of the potentially correlated dependent variables, adds to these concerns. In themselves, each of these alone might not, on further analysis, turn out to be a serious issue. However, taken together, they cast significant doubt on the overall validity of the study.

191. Perhaps the biggest methodological issue is the large numbers of respondents recalling displays well after they have been banned. This suggests the basic measure is capturing a variety of retail attributes, rather than the one desired by the researchers.
192. While this study includes minors, and they are specifically identified in the analyses, smoking initiation as a specific behavior is not measured. The only measures of (self-reported) behavior are of current smoking levels and impulse purchasing. Hence this study does not provide any evidence relevant to smoking initiation in minors.
193. After ban decreases in levels of brand awareness, normative beliefs and perceived impact of displays are better thought of as factors potentially influencing behavior, rather than measures of actual smoking cessation. Smoking cessation itself is not measured and while the observed decreases in current smoking levels from 15 percent to 11 percent 24 months after the ban might provide indirect evidence that some individuals have quit smoking the overall effect here is only weakly significant ($p < 0.05$) and small in magnitude. The authors do report stronger effects for those making a moderate number of store visits, males and those older than 16 years ($p < 0.001$) but again given the various methodological limitations mentioned above and the fact that this is not a direct measure of cessation it is hard to infer how strong this indirect evidence is. Hence this study does not provide reliable evidence on smoking cessation in minors and adults.

Sally Haw, Amanda Amos, Douglas Eadie, John Frank, Laura MacDonald, Anne Marie MacKintosh, Andy McGregor, Martine Miller, Jamie Pearce, Clare Sharp, Martine Stead, Catherine Tisch and Winfried van der Sluijs. Determining the impact of smoking point of sale legislation among youth (display) study: a protocol for an evaluation of public health policy. *BMC Public Health* 14.1 (2014): 1. (Haw et al., 2014).

Winfried van der Sluijs, Farhana Haseen, Martine Miller, Andy McGregor, Clare Sharp, Amanda Amos, Catherine Best, Martine Stead, Douglas Eadie, James Pearce, John Frank and Sally Haw. 'It looks like an adult sweet shop': point-of-sale tobacco display exposure and brand awareness in Scottish secondary school students. *Nicotine & Tobacco Research* (2016): ntw032. (van der Sluijs et al., 2016).

I group these two studies together because van der Sluijs et al. reports the first results from the program of research proposed in Haw et al., 2014.

Haw et al., 2014.

Objectives, background and theory

195. The stated objective of the study is to propose a protocol for examining the impact of the Display Ban in Scotland, which was introduced in large stores from April 2013 and in all other retail outlets from April 2015. The proposed protocol includes a multi-modal, before and after design and mixed data collection methods, which are to be administered over four years. At the time of publication, most of these data have not been collected, so this article is interesting mainly in how the authors approach the challenging design issues in conducting field research on the introduction of a Display Ban. Subsequently I review van der Sluijs et al., 2016 a paper that reports results from the first survey in this program of research. This was conducted in February and March 2013, immediately before the large store Display Ban.
196. The authors' introduction cites several of the previous studies which claim that tobacco advertising and marketing have a direct impact on smoking intentions, perceived prevalence of smoking and youth smoking itself. They also review the history of Display Bans in other countries and some of the limitations of previous research in terms of small samples and short-time periods of observation.
197. The authors next go straight into seven specific research questions they wish to address with their research protocol and eventual data. Summarized, these are:
 - (a) Does the Display Ban result in changes in exposure to tobacco advertising in young people?
 - (b) Is this reduction in POS exposure associated with changes in brand awareness; perceived accessibility of tobacco; perceived prevalence of youth smoking; susceptibility to smoking; and the incidence and prevalence of smoking in young people?

- (c) Is there any evidence of socio-economic patterning in any of the attitudinal or behavioral outcomes in young people?
 - (d) What is the association between area-level deprivation, levels of tobacco advertising and the availability of cigarettes before and after the ban, and the enforcement of the legislation?
 - (e) Is there any evidence of a dose-response relationship between changes in exposure to POS advertising and interim and longer-term outcome measures in young people?
 - (f) Are there unintended consequences of the legislation, for example an increase in black market cigarettes?
 - (g) Is there any evidence of changes in POS advertising and marketing strategies in the lead up to the Display Ban?
198. In my view, answering the first question might raise some issues given the Display Ban was already partially implemented in Scotland at the time of publication. The authors mention completing a baseline study between February and April 2013, i.e. just before the Display Ban came into effect in large stores, and also carrying out marketing audits in May 2013. Some of these data are analyzed and reported in van der Sluijs et al., 2016. In my opinion, it might have been preferable to conduct this baseline study well before the implementation of the Display Ban in large shops, as there are potential confounding factors in the media environment at the time of the introduction of ban. This objection does not apply to the smaller retail outlets as that ban came into effect in 2015; well after the authors had collected their baseline data.
199. In general, the authors identify appropriate research questions and show a broad focus on many potentially relevant factors, including the socio-economic environment as well as the mechanisms by which the various outcomes may be impacted. However, I would suggest that a potential limitation is the lack of a theory or model that ties all these factors together. For example, I might make the argument that environmental, social and individual factors are closely intertwined and that a more comprehensive model would therefore help sharpen any hypotheses about the impact of the ban.

Appropriateness of research design

200. The authors focus on the catchment areas of four secondary schools in Scotland, selected to cover two levels of urbanization and two of socioeconomic status. Thus there is an interesting and atypical emphasis on geography in this study as compared to many others and I would suggest this is a useful methodological innovation. In one important sense the interaction of retail marketing activities and sociological phenomena are “local” to catchment areas and it is therefore valuable to have such a focus in some studies of this topic. For the purposes of this study, the authors define “young people” as aged those aged between 12 to 17 years.

201. **Multimodal research.** The design proposed in this study is a multi-modal one and is made up of the following four components:
1. A visual analysis of tobacco outlets around the schools;
 2. Marketing audits and interviews with a panel of retailers;
 3. School surveys with embedded cohorts; and
 4. Focus groups with pupils.
202. Again, I consider this to be a strong point of the proposed study in that the four components potentially allow them to tie various factors together.
203. **Frequency of proposed data collection is annual.** The authors propose to collect these data annually for four years. They chose a panel of 24 retailers from areas matched with their focal ones but not retailers from the focal areas themselves for this component. As they note, it is important to minimize the chances of biasing the focal studies by drawing attention to the research; they exclude interviews with large supermarkets for the same reasons. On the other hand they conduct a “discrete” audit of the retail outlets in the focal areas, using experienced observers who visited outlets in pairs and recorded key data on displays unobtrusively.
204. In respect of the annual school surveys, the authors introduce an informal hypothesis, namely that the partial (large store) ban will have limited impact. For reasons they do not fully explain they conducted their baseline study only on two age cohorts of students. In contrast, they propose to study a full set of six cohorts in the surveys they will conduct after the introduction of the Display Ban in all shops (i.e. all age cohorts 12 to 17). While the authors believe they can compare their baseline survey to studies from other countries, it is in my view a potential limitation that this baseline is limited to two cohorts, and conducted close to partial ban implementation. Again, it might have been preferable to have a “clean” baseline on all six cohorts well in advance of the partial ban. I also note the authors propose to conduct a second baseline study February to March 2015 ahead of the introduction of the ban in all shops in April 2015 and similar comments might apply to this as well, though one might expect the media environment to be less of a confounding factor at this second stage.
205. Finally, the authors conducted 16 focus groups with their two 13 and 15 year old cohorts in March 2013. This was just after their surveys, allowing them to follow up on issues from these surveys and also eliminating the chances of the focus groups influencing survey results.
206. **Design limitations and proposed ways to overcome them.** The authors note their major methodological challenge was that the Display Ban is introduced simultaneously across all of Scotland at the same time, making it impossible to randomize subjects to experimental conditions or identify a control group. In an attempt to minimize the threats to internal validity and causal influence that their “uncontrolled” design presents they propose various mitigating procedures. First, they include measures of other (non-smoking related) outcomes in their school surveys which, being arguably not impacted by the

display ban, provide an “internal” control on the results. Second, they look at whether there is a dose-response relationship between reduced POS advertising and short, medium and longer-term outcome measures. This can provide some stronger indication of causality. Third, they propose to “triangulate” across their multiple data sources and give greatest weight to outcomes confirmed across all of these. Finally, they propose to monitor other potentially confounding factors in the environment both locally and nationally.

207. The authors show commendable attention to the key methodological issues their design raises. One exception is perhaps the need for a more detailed discussion of respondent recall issues (e.g. as in Carter et al. 2015). It is not clear how these issues will be handled in the survey work and I also make some comments below in paragraph 224 on the limitations I see in some of the measures used by the first survey in this program as reported by van der Sluijs et al. (2016).
208. **Representativeness of sample.** The authors state that they purposely focus on four non-denominational schools with no large immigrant groups, and located in communities reflecting two levels of urbanization and two levels of socioeconomic deprivation. This selection strategy they consider minimizes the impact of any other school factors and gives them adequate coverage of urbanization and deprivation. The other data collection activities follow from the catchment areas of these schools.
209. The authors discuss their sample size and power calculations, which is an unusual and commendable step, especially before data collection. They use the example of Ireland to establish a benchmark of what changes might be expected (McNeil et al. 2010) and they recognize the problem of repeated statistical tests capitalizing on chance events. However, they do use the standard 0.01 and 0.05 probability levels in most of their calculations, whereas higher levels might be preferable. Their calculations assume high response rates but these are typically achieved in school surveys and they correctly acknowledge their focus groups as purposive sampling.
210. Overall, these choices are appropriate though the small number of schools/communities perhaps limits the generalizability of their results to Scotland as a whole. On the other hand, this allows them to focus on distinct geographies and this is a strong point of their proposal.

Measurement issues

211. The authors put forward a “logic” model for their outcome measures (Figure 1 of the study). This links the impact of the Display Ban to short, medium and long-term outcomes (defined as within three months, less than one year, or greater than one year). While this is not a theory as such, it is a useful idea that others in the field could use and develop. It shows a sequence of how these outcome measures should evolve over time and they see this as a way to establish cause and effect more clearly.
212. While the authors list all their various measures they do not provide much by way of details on the specific questions or construct measures themselves.

Appropriateness of analysis and key study results

213. Other than describing their various data transcription and recording methods, and indicating they will synthesize their data into mixed methods matrices, the authors do not discuss how the data will be analyzed by statistical methods.

Overall conclusions

214. In my view, this proposed protocol makes good use of multimodal methods, space and time measures and pre-survey power analysis. Moreover, it attempts to control for internal validity and establish causality through the Figure 1 sequence model. In those senses, this is a model for a good study when, as here, it is not possible to do a more controlled experiment. And in particular, the addition of the geographic, catchment area dimension is an important methodological innovation in this research field.
215. However, I have three reservations that raise concerns about any reliance on this study as a means of demonstrating the impact of a Display Ban. First, whether the authors really have a baseline survey given that large retailers were subject to the Display Ban very soon after the first two cohorts of students were surveyed. There are potentially confounding factors in the environment and also issues of how this baseline is compared to the other four age cohorts. Second, from a scientific point of view, it would have been better to publish the results of this baseline survey prior to the follow-ups. This would be in keeping with the authors' commendable emphasis on setting out their design principles before conducting their research, rather than post hoc rationalizing the design after completion. Third, there is inadequate discussion of the recall and other potential biases in the survey instruments and given the annual nature of these surveys this is an important issue, as raised by others in the field (Carter et al. 2015).
216. As the authors simply propose a protocol for a program of research but do not report on empirical data, this study itself does not provide any evidence against the objectives of my report. I next look at the first published study from this program, to assess whether this provides such evidence.

van der Sluijs et al., 2016.

Objectives, background and theory

217. This study reports results from the first baseline survey of four schools and two age cohorts conducted immediately before the Display Ban in large stores. The authors state their objective as “[T]his study explored the relationship between PoS displays and brand awareness among secondary school students in Scotland” (p.3).
218. The authors discuss their view on the role of POS displays, and cite previous research that such displays are associated with smoking susceptibility and smoking among young people, including several of the studies I have reviewed or cited in my report (e.g. MacIntosh et al. 2012; Spanopoulos et al. 2014 and Bogdanovica et al. 2015). However, this discussion is brief and no theoretical framework or formal hypotheses are advanced. The authors also cite MacIntosh et al. (2012) to the effect that Display Bans in large stores are

likely to have more limited impact on adolescents than those in small retailers. This is perhaps an informal hypothesis.

Appropriateness of research design

219. The authors correctly note that POS displays are not the only influence on adolescent smoking behavior so they include measures of parental and friend smoking behaviors and also involvement in social activities. I would again note there is no mention of the media environment surrounding the introduction of the Display Ban and also no discussion of the “*non-smoking related*” controls outlined in Haw et al. (2014). One strong point of this study is the broader definition of influences on smoking behavior; however, in my opinion, the potential set of controls for confounding factors still remains incomplete.
220. The other point to make on the research design is that the survey is conducted before the Display Ban comes into force. It therefore cannot provide direct evidence on any impact of such a ban, only an indirect case for or against such impact. Furthermore, since the research is conducted at one point in time, causality is difficult to observe, and hence the study can only establish the correlational associations between various factors measured in the survey. The authors acknowledge this point in their subsequent discussion.
221. The authors also conducted sixteen focus groups. While this methodology may be useful for generating questions or ideas it cannot provide tests of any association between factors. Further any explanatory “insights” that might emerge from this methodology are potentially subject to a range of social, group and facilitator biases. I therefore do not discuss the results from these focus groups here.

Representativeness of sample

222. The survey follows the four-school design in Haw et al. (2014) and achieved a response rate of 87 percent, with a final sample size of 1,482 students. This is an excellent response rate so we can largely rule out any non-response bias in this sample. The sample size is also more than adequate for statistical analyses.
223. It should be noted the sample is of S2 and S4 students and so it not representative of all secondary school students. Furthermore, as I noted in my review of Haw et al. (2014), it is not clear whether the data from these four schools can be generalized to the Scottish or UK secondary school population. The authors partly acknowledge this point but argue their demographic measures control for this. This argument is not clear to me as students in other age cohorts are omitted from the survey; therefore it is difficult to see how you can generalize to other age groups.

Measurement issues

224. In my opinion, a potential limitation of the study is the use of prompted measures of brand awareness. That is, students are asked which brands they recognize from a list of 16 brands they are shown, including one fake brand. It is true the authors eliminate students who “recognize” this fake brand from

their analyses and that is a useful step, albeit the numbers of such students are small. However, in my opinion the use of prompted measures is still potentially leading the remaining respondents to an answer. Indeed, standard market research practice is to ask for unaided brand awareness first (i.e. “*which brands do you recall?*”) and prompted brand awareness (i.e. “*which brands on this list do you recognize?*”) at a second step. The contrast between the unaided and aided measures providing a better indication of true brand awareness and potential “*yea-saying*” bias.

225. Other potential limitations concern the measures of exposure to POS displays. There are two of these: frequency of stop visits and “*having seen tobacco products displayed in the last 30 days*” (p. 8). Both were measured separately for large supermarkets and smaller shops. Marketing professionals might find the display question potentially poorly defined for the respondent; from a social science perspective the authors provide no discussion or evidence of the validity of these or any other measures. I would also note that both of these are recall measures and are similar to those that have been criticized by others in this field as open to bias (Carter et al. 2015).
226. Other measures include “*ever smoking*” and “*current smoking*”. However, the latter was not used because of the low smoking prevalence observed in this sample (5 percent of students). Finally, parental, sibling and friends smoking behavior, leisure activities and socioeconomic circumstances were measured through additional questions.
227. Overall, and in my opinion, measurement is a serious weakness of this study, in the sense that no evidence is provided on the validity of the measures used, nor any discussion of potential biases in the responses which, given the nature of the topic studied, could also be a serious issue.

Appropriateness of analysis and key study results

228. The authors use a general linear model to study the association between brand awareness, frequency of store visits and noticing POS displays. The other measures are used as controls so as to identify this association separately from any confounding factors. The joint effect (interaction) of frequency of store visit and noticing displays was not significant and so the authors excluded this from further analysis.
229. Because of the nature of the dependent variable, brand awareness (which the authors categorize into four levels according to number of brands recognized), they use a negative binomial formulation, which is appropriate, and they estimate regression models with and without the “confounding” variables. Here I focus on the latter as providing a better estimate of any effect of exposure on brand awareness. The authors use “relative rate ratio” as their statistic of interest, which is another term for relative risk ratio. That is, the difference in levels of brand awareness between the exposed and unexposed groups.
230. **Key study results.** The regression model with confounders shows a significant association between visiting small shops more than twice per week and brand awareness (relative rate ratio 1.2). The model also shows that

noticing displays in small shops and supermarkets is also significantly associated with brand awareness (relative rate ratios of 1.2 and 1.2 respectively). “Ever smoking” is also significantly associated with brand awareness, as are siblings who smoke, being older and going out with friends three or more times per week. Socioeconomic status, parents who smoke and hanging round the street/park were not significantly associated with brand awareness.

231. I note the sample size for this model is 1,092 students. This is lower than the 1,482 students quoted earlier in the paper, partly due to eliminating those reporting the “fake brand” (76) but possibly also because of missing responses on other questions. This is not discussed in the paper but is a potential source of bias in the regression results.
232. **Effect sizes.** Equating the relative risk ratio to the odds ratio (see discussion in paragraph 109), the effects for frequency of visiting stores and noticing displays in small shops and supermarkets are of small magnitude. Indeed the lower boundaries of the 95 percent confidence intervals for these ratios almost include 1 (no effect). (These lower boundaries are respectively 1.01, 1.03 and 1.01). Hence, if we were to take the stricter 99 percent confidence interval it is possible this would show that these effects are not statistically significant. The strongest effect in the regression model is that of ever smoking on brand awareness (relative risk ratio of 1.5). It is possible this effect would be statistically significant at the higher level of 99 percent, but the effect size itself remains small in magnitude.

Overall conclusions

233. This study has a broad perspective on possible sources of influence on smoking, an excellent response rate to the survey and a large sample size. The analysis also uses appropriate regression modeling methods in a careful and systematic manner.
234. The results of this regression modeling show associations between brand awareness and exposure to POS displays which are of small magnitude (or possibly non-existent if stricter tests of significance are applied).
235. However, I think the more serious issue is the lack of any evidence on the reliability and validity of the recall measures, especially those for the brand awareness and exposure to POS displays. In my opinion, the lack of such evidence calls into question the robustness of the results shown here and any conclusions drawn from them.
236. The study also does not provide any evidence against the objectives of my report partly because it is done before the Display Ban came into effect but more importantly because ‘brand awareness’ is not a behavior as such. It is potentially a precursor to behavior but there are many intervening factors between awareness and action. Indeed on this last point, it is interesting that 79 percent of the students in this sample have never smoked, the vast majority of whom visit stores more than once per week and nearly always notice displays in them (Tables 1 & 2). They are therefore ‘aware of’ cigarette brands

but this awareness does not necessarily result into purchase, and this remains untested in this study.

Ian J. Irvine and Van Hai Nguyen. Retail Tobacco Display Bans. *Forum for Health Economics and Policy* (Vol. 17, No. 2, pp. 169-195). (Irvine and Nguyen, 2014).

Objectives, background and theory

238. The authors discuss that most studies related to a Display Ban are “attitudinal” in nature, and that there are few econometric “event studies” of the sort typically used in the field of economics (and I would add also in the field of finance). Event studies provide a formal quantitative methodology especially designed to identify the impact of a discrete event on an on-going time series of data, for example the impact of a political crisis on economic growth or the impact of a merger on firm value. The objective of Irvine and Nguyen’s study is to undertake such an event study examining the impact of Display Bans and using annual CTUMS data. As an aside, I find it interesting that other authors studying Display Bans have not considered methodology of this type. This methodology could be applied to attitudinal and other data, provided there is a survey repeated over time. Irvine and Nguyen’s use of the term “attitudinal” appears to be restricted to how people feel about bans and assess their likely impact, whereas other behavioral scientists might have a broader definition and survey data that could be equally amenable to event studies. That said, the authors point out that some courts have not given data based on people’s opinions much weight, perhaps because of the low correlation between opinions and behavior (Huang & Fong, 2013)²³ and therefore their focus is on *behaviors*.
239. The authors note that the CTUMS data are available from 1999 and involve large samples of monthly interviews, allowing precise before/after analyses. The 10 Canadian provinces also implemented Display Bans at different times between 2003 and 2009, which is important because it provides them with multiple events to study and thus adequate variation and degrees of freedom for their statistical analyses. Interestingly, they state that data like these are not available in many countries; to have multiple events you need a federal structure such as Canada with different times of implementation of the bans, and for precision in estimating effects large samples are also needed, implying significant investment in data collection. I would note that the sequence of events in Canada provides natural control groups.
240. The authors then review existing work relevant to the issue of a Display Ban. Rather critically, their view is that most of the studies are opinion-based and are subject to many limitations. For example, exit-interview based studies such as Clattenberg et al. (2013) or Wakefield et al. (2008) “*do not investigate if a smoker, not having purchased cigarettes as a result of the absence of displays, may purchase cigarettes at another retail outlet before exhausting his or her stock*” (p. 173). Equally, the authors consider that many studies suffer from problems of statistical inference, mentioning endogeneity (here they cite difficulties in establishing cause and effect), poor measurement (such

²³ Huang, F. C. & Fong, G. (2013). Cigarette Graphic Warning Labels and Smoking Prevalence in Canada: A Critical Examination and Reformulation of the FDA Regulatory Impact Analysis. *Tobacco Control*, 1–6. doi:10.1136/tobaccocontrol-2013-051170.

as failure to distinguish between the number of exposures to advertising and the intensity of each exposure), and “unobservables” (such as whether teenagers who smoke are more likely to be the sort of personality who “hang outs” in corner stores, confounding measures of exposure in those stores with personality). In summary, the authors suggest that many of the existing studies on the topic of Display Bans might be explained by factors other than the ones the authors of these studies themselves considered.

241. Finally, the authors point out that while there are a number of econometric studies of smoking behaviors they were not able to find any on display bans as such (as of 2014).
242. In conclusion, the authors do an excellent job of pointing out many of the limitations of opinion-based studies in this area and they make an important contribution by introducing the event study methodology to Display Ban studies in the peer-reviewed literature. However, I would note with irony that their survey data would also be, in the eyes of some behavioral scientists, “opinion-based” as it is self-report of what people think they did rather than independent observation or tracking of behavior. In their defense, one could argue that self-report of “what I did recently” is a more reliable measure than “what I did in the past” or “what I intend to do in the future” and there is support for that conclusion in the consumer research literature.
243. In summary, the authors do not introduce a theoretical framework for the impact of Display Bans, their focus is more on innovation in methodology.

Appropriateness of the research design

244. The authors use a differences-in-differences methodology. i.e. they look at changes before and after the Display Ban but they subtract the changes observed in a control group from those observed in the treatment group. Effectively this reduces the impact of unobserved confounding factors in the environment, which are assumed to have impacted both the control and treatment groups equally. It also controls for overall trends in smoking outcomes that are not related to the ban. This methodology is often used in medical and policy studies, where a treatment is applied to a subset of the population (Athey & Imben, 2006).²⁴ Here, the authors use the data from before the introduction of a Display Ban as the control. This is, they admit, not as strong as having a separate control group post-ban, but that is not possible in this circumstance, and what they can do does go some way to controlling possible confounds. They also include individual level and province level covariates (“control variables”) accounting for the heterogeneity of any other possibly confounding effects. Finally, they estimate linear and quadratic time trends. The authors do not discuss this methodology at length as it is well established in the literature, but they do provide citations to several of the standard references and key papers on this differences-in-differences methodology. Standard econometric time-series regression techniques are then applied to these difference-in-differences.

²⁴ Athey, S. & Imbens, G. (2006). Identification and inference in nonlinear difference-in-difference models. *Econometrica*, 74(2), 431-498.

245. The authors spend some time discussing the stated problem of contraband products, which they regard as an important problem because illegal sales were at a peak during the time display bans were introduced in Quebec and Ontario. Significant sales of contraband products may distort official price data and people buying them may also be less truthful in surveys. Hence, the authors suggest that any regression estimates they obtain may be biased. More critically, this makes it difficult to establish causality between Display Bans and cigarette sales in these two provinces. Their “primary strategy” to ameliorate this problem is to run analyses for all available years on the other 8 provinces i.e. excluding Quebec and Ontario. A second approach is to run analyses on all 10 provinces but only before contraband begins to emerge as a problem, which they say is sometime during 2005. A weakness they recognize in their second approach is that only two ban “events” have occurred by 2005, weakening the ability of their regression models to identify ban effects.
246. Overall, the authors demonstrate a good appreciation of the potential problems in their design and analysis, but the necessity of omitting of two major provinces because of the contraband problem is clearly a limitation on the generalizability of their results. On the positive side their discussion of the contraband issue shows the importance of considering alternative explanations for any observed decline in smoking that occurs at the same time as a Display Ban.

Representativeness of the sample

247. The authors briefly discuss the content of the CTUMS surveys which they believe have rich information on smoking behaviors beyond usage, including smoking intensity and quit intentions (though use of the latter might open them to criticisms of using more opinion-based measures). They do not actually talk much about the survey method, for example coverage of the population, how the data is collected, response rates, or any issues of over or under-sampling, etc. This is presumably because CTUMS is a publicly available and well-documented survey conducted using established survey practices. However, significant levels of non-response might raise concerns about the generalizability of the survey results to the Canadian population as a whole.

Measurement issues

248. The authors focus on four main outcome variables from the CTUMS surveys; (1) prevalence of smoking, (2) number of cigarettes smoked by week, (3) intentions to quit and (4), for some robustness tests, quit attempts. As noted above the third of these is not really “behavior” in the typical understanding of the term and indeed the authors place less weight on their analyses of this outcome. The other outcomes are measures of self-reported behavior.
249. The authors discuss that these surveys do not distinguish between new and existing smokers or “quitters”, so the prevalence measure is essentially a “net” indicator of any impact of a Display Ban. To compensate, they also look at inflows and outflows from the stock of cigarettes smoked via the number of cigarettes smoked per week, arguing that the impact of a ban on the amount of

tobacco intake is “*possibly as important*” as its impact on prevalence of smoking.

250. Again, the authors do not discuss the nature of the survey questions themselves, which is a limitation as others have raised measurement issues around survey questions (for example, Carter et al., 2015). On the other hand, the CTUMS surveys were monthly so recall bias may be less of an issue in these data. Irvine and Nguyen do suggest that smokers typically under-report their usage, but they do not consider this an important bias because the difference-in-differences methodology focuses on changes not absolute levels.

Appropriateness of analysis and key study results

251. The authors start by presenting simple de-seasonalized trends in smoking prevalence and weekly quantity of cigarettes smoked by province. These do not indicate big changes around the time of the various Display Bans being introduced, suggesting ban effects, if any, are small.
252. They then apply standard regressions applied to the difference-in-differences data. Here there is not much discussion of regression modelling issues, possibly because the approach is well known, but it would have been helpful to include some discussion of the applicability of these techniques to the form and nature of the CTUMS outcome measures.
253. **Key study results.** In respect of the regressions on the prevalence measure, the regression results show no significant impact of the Display Bans even with a small coefficient (close to zero). Their various controls do have impact i.e. age (an inverted U relationship), education (less likely to smoke), married (less likely to smoke), English/French speakers (more likely to smoke), men (more likely to smoke) and large households (less likely to smoke). The trends in smoking prevalence over time are downward. The authors, as economists, note that the observed zero impact of price is hard to understand, but possibly this is because prices are highly correlated with the overall time trend, as they show with a subsequent test of collinearity.
254. In respect of regressions on the smoking intensity measure (where the authors use the log of number of cigarettes smoked by week because of the skewed nature of the distribution and data for smokers only), they basically obtain the same results as for prevalence. In particular, there is no statistically significant effect of the Display Ban, and their various controls show similar effects to the prevalence analysis. However, for intensity, price behaves in the expected way and has a significant negative coefficient, indicating that higher prices lead to less smoking intensity.
255. The regressions on the intention to quit measure show similar results to the two behavioral outcomes.
256. The authors also conduct an extensive series of what they call robustness tests. First, they redo their prevalence analyses using probit regression instead of the more standard ordinary least squares regression. This reanalysis would go to the issue I raised above on the nature of the survey measures. Probit regression makes fewer assumptions about the prevalence measure. They find this makes

no difference to any of their results, except when they include a quadratic time trend, in which case the display ban coefficient is weakly significant at the 5 percent level but has the wrong sign (i.e. increases smoking prevalence). Second, they use “quit attempts” rather than “intentions to quit”. Again this makes no difference to their results in that the ban did not lead to an increase in quit attempts. Here I would again note that “quit attempts” is a (self-reported) behavior, whereas “intentions to quit” is not. Third, and as described above, they exclude the period of smuggling, again with no difference to their results; the Display Ban coefficients are small and non-significant. Finally, they control for the public smoking bans that also came into play during this period. Again they find including this additional control makes no difference to the results

257. Further, the authors also look at whether, while they observe no effects of Display Bans on the total population, there might be effects on subgroups of this population. To do so they repeat all their analyses by separate age groups. Nearly all of their analyses result in the same conclusion—no effect of a Display Ban. However, there is one negative and significant coefficient for the youngest age group (15-24) on smoking intensity (log of the number of cigarettes smoked), albeit that this only appears at the $p < 0.05$ level and when a quadratic time trend is included. On deeper analysis of legal smoking ages, they conclude the Display Bans “may” have impacted the intensity of legal youth smoking consumption, but not smoking prevalence or quit intentions. Here, I would note that the authors have very large sample sizes (many tens of thousands of data points) and I would have consequently preferred tests at a higher level of significance, in which case these latter effects may well not be significant. Finally, the authors look at whether there may be differences between occasional and daily smokers, finding no impact of bans on either group.
258. In their conclusions, the authors raise the issue of lags, that is, whether the effects of Display Bans are immediate or emerge over longer periods of time. They do not consider lags in their analyses, a choice they argue is right because (1) there is no theory on what these lags should be and so any lagged analysis becomes an exercise in data mining, (2) if displays are a memory trigger the impact should be immediate and (3) the majority of smokers at any time intend to quit so again you might expect the ban to have immediate effect by tipping some to actually quit. They also point out that their dummy variable for the Display Ban essentially measures the average impact over all post-ban periods, a “switch” they turn on in the month after the event and which provides a test of a difference in behavior before and after the ban. Their carefully stated conclusion on this section is that they “*do not claim to have ruled out some longer term impact*” (page 193) but that they do not detect a change in behavior due to the introduction of Display Bans.
259. **Effect size/behavior.** Most analyses show no significant associations so it is not appropriate to discuss effect size for these. Insufficient information is presented in the paper to estimate the magnitude of the one result on smoking intensity for 15-24 year olds. The authors’ discussion in the text seems to imply this is a not a strong effect and potentially an artifact of the time trend

assumption (my interpretation) since the effect is not significant when alternative assumptions are tested.

Overall conclusions

260. In my view, this is an excellent econometric study with great attention to potential confounding factors in the data and equal care on what can and cannot be stated from their data, analyses and results. The difference-in-differences and event study approaches bring something new to the topic and the many robustness tests and subgroup analyses lend greater confidence to their results and conclusions. In most of their results they find “*no systematic support for a significant impact of the bans on participation or quit intentions, and just limited support (among youth) for a reduction in intensity*” (page 192).
261. The limitations of the study therefore lie less in the author’s techniques and analyses than in the basic self-report survey measures contained in the CTUMS data, and the fact that two major provinces have to be omitted from the primary analyses because of the acknowledged problem of contraband cigarettes. The former raises some issues of potential response biases, though the difference-in-differences approach goes some way to controlling for this, while the latter limits the generalizability of the study to the total Canadian population.
262. Turning to the objectives set out in paragraph 6(a) and (b) above, smoking initiation among minors is not directly addressed by this study. As the authors point out prevalence is a net measure of initiation and quitting and therefore “*youth initiation impacts are only captured implicitly*” (page 179). Furthermore, they find no significant association between the introduction of a Display Ban and smoking prevalence, either in their overall data or in their age subgroup analyses.
263. In respect of smoking cessation in minors and adults, quit intentions are not the same as quitting, and further the authors find no significant associations between the introduction of a Display Ban and quit intentions, either in their overall data or in their age subgroup analyses. In their robustness tests, the authors also look at quit attempts, but again they find no significant associations. What they do find is limited support (i.e. a weak effect) for some reduction in consumption amongst those in the 15 to 24 age groups.
264. I would note that, in preparing this report, I identified two further public available econometric studies utilizing the same data set as Irvine and Nyugen: (Lilico, 2009 and Padilla, 2010).²⁵ All three reports reach a similar conclusion, namely that a Display Ban has had no statistically significant impact on smoking prevalence in Canada. These three different reports use different analytical approaches to the CTUMS data, but reach essentially the same

²⁵ Lilico, A. (2009). The impacts of restrictions on the display of tobacco products. *Europe Economics*.
Jorge Padilla. (2010). The effectiveness of display bans: the case of Canada. *LECG*.

conclusion. The same point can be made about the different model formulations tested by these three studies, which have basic similarities but which also, across the three, cover an extensive range of modeling possibilities.

265. In addition, and along with Irvine and Nguyen, Lilico (2009) and Padilla (2010) highlight the role of price. According to these studies, price changes are likely to have a meaningful impact on the prevalence of cigarette smoking and the number smoked by day, and studies that ignore price may risk confounding the impact of regulatory changes with price trends in the market.

Annice E. Kim, James M. Nonnemaker, Brett R. Loomis, Asma Baig, Edward Hill, John W. Holloway, Matthew C. Farrelly and Paul R. Shafer. The Influence of tobacco displays and ads on youth: a virtual store experiment. *Pediatrics* 131.1 (2013): e88-e95. (Kim et al., 2013)

Objectives, background and theory

266. The objective of this study is to conduct a virtual experiment to examine the potential impact of banning displays and advertising at POS on youth outcomes. Youth they define as 13-17 year olds and outcomes as (1) perceived ease of purchasing tobacco in the virtual store and (2) purchasing attempts at the virtual checkout.
267. The authors cite prior studies that show young people are highly aware of POS marketing and which indicate that those who are exposed to such marketing are more likely to experiment and to become smokers. In particular, they discuss the Wakefield et al (2006) experiment where young Australians were asked for their reactions to photographs of cigarette packs with and without display advertising. However, while this study suggests that those exposed to such advertising perceived that they would be less likely to be asked for proof of age and that more stores would sell them cigarettes, it is limited by conducting the experiment in the classroom rather than a “real world” retail environment. So the contribution of these authors is to extend Wakefield et al’s work to a virtual store environment that is closer to this real world environment. The authors note that virtual experiments have been used in smoking research before (e.g. Traylor et al, 2008)²⁶ but not, to date of publication, on tobacco displays and retail store advertising.
268. The study provides only a short introduction with a small number of citations to prior research and little theory as such. Possibly because of this, no formal hypotheses are advanced, and the authors state their primary aim is to “assess” whether youth exposed to a display ban condition are less likely to attempt to buy at the virtual store and also perceive that it is difficult to buy cigarettes at this store. A secondary aim is to “examine” whether the presence of tobacco advertisements in the store moderates these relationships.

Appropriateness of the research design

269. The authors devote a significant amount of their paper to describing their innovation, the virtual store experiment. Their experiment itself has a simple three by two design and therefore six conditions in total. That is, three display conditions; (i) open displays; (ii) enclosed displays; and (iii) enclosed displays with advertising on the display cabinet, together with two advertising conditions: advertising in-store or no advertising in store. Though the authors do not mention this, their design is a between-groups experiment, where participants are randomly assigned to one of the six conditions.

²⁶ Traylor, A.C., Bordnick, P.S. & Carter, B.L. (2008). Assessing craving in young adult smokers using virtual reality. *American Journal of Addiction*, 17(5), 436-440.

270. They used an off-the shelf virtual convenience store which they adapted and then tested with a group of 12 people aged 14 to 17. This led to further improvements in terms of faster navigation and expanded product selection. The authors do not discuss why they chose this experimental design (as opposed to other possibilities), though a footnote states that one of the conditions is modeled after New York regulations at the time.
271. The online task for participants in the experiment was to select four items to buy from the store, namely: a snack from the aisles, a drink from the coolers and two items of choice from the checkout. They had 10 minutes to do this, though most took far less time, an issue I will return to later. If the participant clicked on the tobacco display in the store (in conditions where this is present), the virtual cashier then asked whether they want to buy tobacco. If they said “yes”, the cashier told them they were too young; if they said “no” the cashier told them “OK”. All participants were then exposed to the display for another 10 seconds before being sent to a survey page to collect their post-experiment perceptions and opinions.
272. Overall, and in my opinion, this seems to be a simple, well-designed experiment where participants would not necessarily know the purpose of the study until they got to the survey (which appears to be totally focused on cigarette purchasing). This implies their behavior in the store is less likely to be biased by any social desirability effects around smoking, which is potentially a strong point of the study. It is true that, as the authors rightly report, they had to mention smoking in the screening procedure and parental consent form. But, as the authors themselves note, such effects should have been consistent across conditions, and, in my opinion, the time delay from recruitment to experiment also lessens this threat.
273. Photographs of the virtual store provided in the paper also suggest the store environment is reasonably realistic. In addition, those participants who click on the display as one of their choices were providing a measure of actual behavior rather than opinion. Moreover, the participants were not asked to purchase cigarettes, just to buy two items at the checkout, along with the items they bought in the aisles, again disguising the intent of the researchers, and potentially reducing social desirability biases.
274. A general point on experiments like this, of which the authors are aware, is that social scientists regard them as typically high on internal validity but lower on external validity. That is, because of the random assignment of participants to conditions, and the use of control groups, we are confident that any statistically significant effects of the treatment (here a Display Ban) we observe are really present and not due to other factors. This would be high *internal* validity. In contrast *external* validity deals with how closely the behavior of people in the experiment resembles their behavior in the real world. This is harder to achieve in experiments because there is an obvious trade-off between control and realism. In my opinion, this study has a good balance between internal and external validity. The authors have the control of an experiment but the virtual store and the task they set within this store has a reasonable approximation to behavior in the real world.

Representativeness of sample

275. The authors recruited their experiment participants from established market research panels with a focus on 13 to 17 year olds. They note that this is a convenience sample and therefore cannot be generalized to the national population. One issue in a between groups design is to ensure that the randomization across conditions results in similar groups, otherwise results may be misleading. Here the authors provide data (Table 2) indicating that the distribution of demographics was similar across the six conditions. Another issue often raised is whether the groups are of adequate size. However, I do not consider this to be a problem here because the overall sample the authors obtain is large by the standards of most experiments.
276. The authors note that 4,189 people consented to do the experiment but only 1,216 completed it. This is a 70 percent drop out for a study that might be considered to be 10 minutes of fun and for which participants also received a US\$6 incentive. The authors do not discuss the reasons behind this drop out, though their Figure 3 shows the numbers who (1) never clicked on the first link, (2) who did not install the necessary browser plugin, (3) did not complete the shopping tasks and (4) did not complete the survey. They also note there are no “meaningful” differences in the profiles of the initial and final sample, but they do not provide details of this analysis. In my view, this is a significant limitation of the study because it raises the possibility that those who completed the experiment are on some unknown dimension different to those who did not (Burton, et al., 2012 also make this point). This could further limit the generalizability of their results.

Measurement issues

277. The authors have two key outcome measures, the perceived ease of purchasing cigarettes in this store and whether the participant tried to purchase cigarettes by clicking on the display. In my opinion, the second measure is the better one because it occurs before the survey heightens the salience of cigarettes. The first, perceptual measure is more suspect, because at that point participants know that the topic is cigarette purchase and (if they clicked on the display) they have also just been reminded (by the virtual cashier) that this is illegal at their age. Hence it is possible social desirability biases confound the perceptual measures. The authors appear to agree with my reservations as they also advance clicking on the display as the more valid measure and suggest (implicitly) that the perceptual measures are of less worth.
278. They also raise the issue that participants only spent 3.4 minutes in the virtual store, which is less than the 10 minutes they anticipated and also below the 16 minutes average that young people spend in real convenience stores. This may suggest a superficial exposure to the experimental conditions, particularly exposure to the display at checkout. The authors argue that it does not, suggesting, correctly in my opinion, that advertising studies often use exposures of a few seconds and therefore the 32 seconds the average respondent spent at the virtual checkout is an adequate exposure, but they do acknowledge this as a potential limitation to their work.

Appropriateness of analysis and key study results

279. The authors use logistic regression because their dependent click measure is binary (clicking/not clicking on the cigarette display). The experimental conditions then become dummy independent variables and they also use a number of covariates measured in the survey (smoking behavior, usual source of cigarettes, social influence to smoke, a sensation-seeking behavior scale and demographics). The omitted reference condition for the experimental condition dummy variables was the current store environment of open display and store ads. This is a standard regression approach and appropriate given the nature of the dependent variable. The covariates also help identify more clearly the true effects of the experimental conditions. Here the authors split their data into two subgroups for analysis, “*current smokers (N=409)*” and “*open to smoking (N=723)*”.
280. **Key study results.** For the current smokers subgroup, their results show significant effects on attempting to purchase cigarettes ($p < 0.01$ for three of five coefficients, $p < 0.05$ for one coefficient), with the implication that all youth who were exposed to any of the enclosed display conditions were less likely to attempt to buy cigarettes. Various comparisons between conditions also show significant effects ($p < 0.05$) consistent with this conclusion. These effects are weaker for their “open to smoking” subgroup. Indeed, “open to smoking” youth in the enclosed display, no advertising condition show no differences in purchase attempts compared with the reference condition of open display and store advertising.
281. The authors describe their results on post-experiment “*perceived ease of access to cigarettes*” as “*inconsistent*”, which is perhaps not surprising given the limitations of this measure, and the coefficients of the logistic regression themselves are not significant, though a few between condition comparisons are ($p < 0.05$).
282. **Effect size/behavior.** The raw comparisons show clear effects, for example 24 percent of participants in the open display condition with advertising clicked to purchase cigarettes compared with only 9 percent in the enclosed display condition with advertising. These numbers fall to 16 percent and 11 percent when no store advertising is present. However, these raw comparisons do not take into account the covariates in the regression models, which provide better estimates of the effect of enclosing the display. Looking at the odds ratios from those regressions for “current smokers” and those “open to smoking” shows several that are statistically significant, ranging in magnitude from 0.39 to 0.22. Which, following paragraph 42 above, indicates *moderate* to *large* magnitude effects on purchase attempts.

Overall conclusions

283. The authors themselves conclude their results show enclosing displays “*could deter youth from attempting to purchase tobacco at retail stores*” (page e94]). However, they also note their primary concern in designing the experiment was to “*minimize threats to internal validity at the cost of potentially limiting external validity*” (page e94). I would argue that internal/external validity is indeed the heart of the matter.

284. On internal validity, the study is well done, with commendable attention to design details. The virtual store is a novel and, I would suggest, potentially better way of collecting data than the typical classroom survey/experiment, and the experiment yields a measure of behavior as opposed to opinion. The authors themselves raise the issue of priming participants (due to the recruitment and consent process) as a potential threat to internal validity. However, I think that randomization, the time between consent and experiment and the fact that the task itself is to buy any four items, reduces this threat. I believe this experiment does have high internal validity.
285. I have concerns on the high drop out during the experiment and the extent to which their results are driven by small subgroups of participants. This goes to external validity. I would therefore restate their conclusions as indicating that enclosing displays *might* deter *some* youth from attempting to purchase tobacco, but that whether this occurs in reality remains to be demonstrated by further research.
286. The other issue of external validity is how realistic a virtual store shopping experience is compared to one in a real store. Here, I am less concerned by the realism of the virtual store experience than how the experimental task relates to normal shopping behavior. In particular, how does this “one off” shopping task compare to behavior in regular visits to the convenience store? Again this suggests the need for further research to establish the external validity of the virtual store approach.
287. In respect of the objectives of a Display Ban at paragraph 6 (a) and (b) above, this study focuses on 13 to 17 year olds. The authors also analyze “open to smoking” participants separately to those who are current smokers. However, the most relevant comparison shows no significant effect of enclosing the display on this “open to smoking” subgroup. Hence, this study provides no evidence to suggest that a Display Ban would have any impact on smoking initiation. The results of the study do suggest that enclosing the display does reduce purchase attempts by minors. Of course, this is not the same as quitting smoking, which is not specifically addressed by this study.

Annic E. Kim, James M. Nonnemaker, Brett R. Loomis, Paul R. Shafer, Asma Shaikh, Edward Hill, John W. Holloway and Matthew C. Farrelly. Influence of point-of-sale tobacco displays and graphic health warning signs on adults: evidence from a virtual store experimental study. *American Journal of Public Health* 104.5 (2014): 888-895. (Kim et al., 2014)

Objectives, background and theory

288. This study employs the same virtual store experimental approach as Kim et al. 2013, which I discussed above, but here with a focus on adults rather than youth and different experimental conditions. In this study, the authors have the same POS display conditions as before but they consider the absence or presence of graphic health warning signs rather than the in-store tobacco advertising conditions in Kim et al. 2013.
289. The authors are cautious in their citations to previous studies, noting that tobacco advertising and displays “may” act as cues to smoke, to stimulate first purchasers or influence relapse among recent quitters. They cite the Wakefield et al. (2008), Carter et al. (2009) and Germain et al. (2010) studies to support some of their propositions.
290. In particular, the authors argue that virtual experiments “may” have better ecological validity than traditional methods, and they cite a number of studies in support of their argument (for example, Hubal et al, 2008 on “embodied conversational agents”²⁷ or Bauman and Sayette, 2006 on “smoking cues in a virtual environment”²⁸). Ecological validity is generally defined as how well the setting of the study approximates the real life setting. Here this would be a matter of how well the virtual store format reflects a real convenience store format. This is not the same as external validity, which is defined as how well the *behavior* of subjects in the study approximates behavior in real life. Here this would be how well the behavior of subjects in a virtual store experiment approximates their behavior in a real store. This is an important distinction because, in general, experiments sacrifice external validity for internal validity, a point discussed in my analysis of their earlier work, and which they also recognize in that work and in the current paper.
291. While the authors provide a nuanced and careful introduction to their topic, and possibly cite a broader literature than others in this field, they advance no theoretical framework or formal hypotheses, which might lead to sharper tests or better explanation of observed phenomena. Rather they set their objective as “*to assess whether exposure to enclosed tobacco display and a graphic health warning sign decreased urges to smoke and tobacco purchase attempts*”. Equally, in the context of this objective, it might also have been

²⁷ Hubal, R.C., Fishbein, D.H., Sheppard, M.S., Paschall, M.J., Eldreth, D.L. & Hyde, C.T. (2008). How do varied populations interact with embodied conversational agents? Findings from inner-city adolescents and prisoners. *Computers in Human Behavior*, 24(3), 1104-1138.

²⁸ Bauman, S.B. & Sayette, M.A. (2006). Smoking cues in a virtual world provoke craving in cigarette smokers. *Psychology of Addictive Behaviors*, 20(4), 484-489.

useful to discuss the merits of experiments versus other methods in slightly more depth.

Appropriateness of research design

292. The focus of the study is current smokers and recent quitters aged 18 years or older as opposed to the 13 to 17 years olds of their previous work (Kim et al. 2013). As the participants are adults, there was no need to mention tobacco in the recruitment procedures and so the authors claim participants were blind to the study conditions, simply being told the study was about consumer shopping behavior. However, in the discussion of the limitations of the study, the authors mention that their screening questions asked about smoking behavior. These two statements are inconsistent. However, as in the previous study, I do not believe this is a major limitation because of the randomization of participants to experimental conditions and the delay between consent and completion of the online experiment.
293. The authors use the same three by two design format as their previous study. The three display conditions are the same as before, namely an open cigarette display, an enclosed display and an enclosed display with tobacco advertising. The other two conditions are different, namely the absence or presence of a graphic health warning at the POS. All conditions also had in-store tobacco advertising and while this would not affect the experimental outcomes per se, it might well raise issues of ecological and external validity, depending on whether any Display Ban also bans in-store advertising of tobacco products. Thus the ecological and external validity of this experiment depends on its alignment with ban conditions in any future legislation. In some countries in-store advertising has been banned before or at the same time as POS displays, implying that the results from this experiment cannot necessarily be applied to other countries. The authors do not discuss this point in depth, except to mention in their closing discussion that the presence of such advertising may have diminished the influence of the graphic health warnings.
294. The online task for participants is also similar to before, namely to buy four items, a snack from the aisles, a drink from the coolers and two items of choice from the checkout. However, when a participant clicks on the cigarette display the virtual cashier asks them “*if they want to buy tobacco products*”. If they say yes, the program then provides them with a list of brands from which to select.

Representativeness of sample

295. Recruitment procedures were similar to Kim et al. (2013) and resulted in 1,216 people completing the shopping task and survey out of 4,189 who were invited to participate. While the authors again provide an excellent summary of their sampling framework in Figure 1, this is also a 71 percent drop out over the course of the exercise, limiting the generalizability of their results. As before, the authors correctly note their sample is a convenience one and cannot necessarily be extrapolated to the population at large.

Measurement issues

296. The authors measure two main outcomes. First, “purchasing” in the sense of clicking on the display and confirming that they want to buy tobacco. The authors suggest that this can be regarded as a behavioral task. I would agree with them that this is behavior, however the question of external validity would still remain in that clicking on the screen is not the same as interacting with a cashier. Second, an urge to smoke measure that is administered through questions in the post-experiment survey. The latter is clearly a self-report and not a behavior, and as they note is prone to more confounding factors and biases (for example, the presence of others in the home when the online task done).
297. In the survey part of the study, the authors also collect unaided recall of products seen in store, an aided but verified response on the health warnings and the extent to which participants thought the convenience store resembled a real convenience store. The latter is a useful addition to their previous work, assessing the ecological validity of the virtual store setting.

Appropriateness of analysis and key study results

298. The authors use ordinary linear regression on their urge to smoke measures and logistic regression on purchase attempts and as in the previous study ran separate analyses on current smokers (N=712) and recent quitters (N=465). To estimate main effects and the interactions of displays with health warnings they combine all the enclosed display conditions into one variable (enclosed versus open) and they include a series of control variables (age, gender, ethnicity, education, frequency of convenience store visits and for smokers and quitters some additional specific smoking behavior controls). In these models they were not able to estimate the effects of the presence of advertising on the displays because this was too highly correlated with the enclosed display variables.
299. **Key study results.** In terms of raw effects, there are significant differences in how many participants saw tobacco products in the open versus enclosed conditions (89 percent versus 48 percent). Participants also saw the health warnings when they were present (37 percent to 53 percent in the presence of health warnings versus 3 to 6 percent mentioning warnings in conditions where they were not actually present). In terms of *purchase attempts*, when warnings were not present, more participants attempted to purchase cigarettes in the open display conditions than the enclosed display conditions (38 percent versus 2 percent). When warnings were present, there was the same effect (from 33 percent to 3 to 5 percent).
300. However, the logistic regression results on *purchase attempts* show a different picture. This is because the regressions are more likely to reveal the real effect of displays after accounting for the many other factors (as listed above) that may influence the raw results. For current smokers, there is a significant effect of enclosing the display, which lowers purchase attempts ($p < 0.01$). However, there is no effect of the health warnings themselves or any interaction of these warnings with enclosing the display. The authors were

unable to replicate this analysis on recent quitters because few quitters, in any condition, attempted a purchase.

301. The linear regression results on urge to smoke are weaker for current smokers. There is a main effect of enclosing displays, which lowers the urge to smoke but at a reduced level of significance ($p < 0.05$). Again, there is no effect of health warnings or the interaction of these warnings with enclosing the display.
302. As is the case in Kim et al. (2013), the authors do not present the coefficients for their covariates, which is a pity as it might have allowed deeper understanding of what drives the raw differences and how important the enclosed display effect is compared with other effects. Also given the reasonably large sample sizes, one might prefer tests at a higher level of statistical significance.
303. The authors conclude that in a virtual store enclosed displays lower both attempts to purchase and the urge to smoke. In contrast, graphic health warnings do not affect either outcome, in their opinion possibly because of limited exposure and/or the presence of in-store advertising. In my opinion, this rather suggests the inclusion of in-store advertising was a mistake as it confuses the results. It should either have been omitted or a more complex design used, for example three display conditions by two health-warning conditions by two in-store advertising conditions (presence/absence).
304. As with their youth experiment, participants here completed the task very quickly, namely in 3.6 minutes.
305. **Effect size/behavior.** For behavioral outcomes there is one statistically significant odds ratio. This is 0.06 for the main effect of enclosed displays on reducing purchase attempts by current smokers. Using my numbers from paragraph 42 earlier, this represents a *large* magnitude effect on purchase attempts.

Overall conclusions

306. As with their youth study, the results of this study suggest that, in a virtual store environment, there is an effect of a Display Ban in lowering purchase attempts for current smokers. However, while the results here show the virtual store setting to have good ecological validity, I would suggest that, as before, the convenience sample and the large drop out during the experiment limit the external validity of the study. Indeed, the authors are careful and correct to state that their results derive from this virtual setting, and only if they translate to a real setting (that is, have external validity) should others draw policy implications from them.
307. The authors also identify as a limitation on external validity, the one time, short exposure to a virtual store, compared with the repeated exposure individuals may have to real convenience stores. However, without a strong theory on exposure/non-exposure and additional research, I would suggest that it is difficult to say what the impact of repeated exposure to enclosed displays would be. It could be that purchase continues to decline or that consumers

adjust and return to their previous level of consumption. Further the drop out during the experiment might possibly indicate that those who completed this virtual task differ on some unknown dimension from the overall population of smokers, further limiting our ability to extrapolate these results to the real world.

308. As far as these results relate to the objectives of my report then this study does not consider minors and so does not address *smoking initiation in minors*. Concerning *smoking cessation in minors and adults*, this study focuses only on adults and does not specifically measure cessation as such, like their previous study. For adults the experiment shows that purchase attempts in a virtual store decline when displays are enclosed. Given the care the authors took in designing their experiment this result has high internal validity so we can be confident the reduction of purchase attempts exists, just as in their previous study of minors.
309. For both studies, the key question is how well the Display Ban effect Kim et al. observe generalizes to real world settings and populations. We know it will generalize to some sub-groups of the population (i.e. those similar to the individuals who eventually took the two experiments) but we do not know how big these sub-groups are and, more importantly, we do not know how well behavior in the virtual task resembles that in a real store visit. Finally, even if the virtual task does resemble the real store, there is the issue I raised above, namely the relationship between a single visit and/or purchase and the ongoing visits and consumption typical of consumer behavior over longer time periods.

Lin Li, Ron Borland, Geoffrey T. Fong, James F. Thrasher, David Hammond and Kenneth M. Cummings. Impact of point-of-sale tobacco display bans: findings from the International Tobacco Control Survey. *Health Education Research* 28.5 (2013): 898-920. (Li et al., 2013)

Objectives, background and theory

310. This study is cited by many of the studies I have reviewed, primarily because it provides a multi-country perspective, using data from Wave 5 (2006) to Wave 8 (2010) of the International Tobacco Control (*ITC*) Survey. This is an annual survey that runs in Australia, Canada, the UK and the USA, albeit I understand that the underlying raw data is not publicly available,²⁹ which is regrettable, as this would allow others to replicate the results.
311. The authors discuss some prior literature and use a review by Henriksen (2012)³⁰ as their motivation for the need to look at the immediate and long-term consequences of POS marketing. They also suggest there is public support for Display Bans and these have been introduced in several countries, citing public reports that their introduction has been associated with a decrease in smoking prevalence (e.g. Health Canada, 2011). However, they acknowledge studies that suggest that any reduction in smoking prevalence resulting from a ban is likely to be gradual (e.g. McNeil, Lewis and Quinn, 2011) rather than immediate and so it is important to monitor this over time. Moreover, they note there are no inter-country comparisons and little published research using such longitudinal methodologies, a gap they aim to fill with this study.
312. The authors therefore state their objectives as examining the variability in POS market restrictions in the four countries and the effect of these varying restrictions on adult smokers' exposure to marketing and cigarette purchasing.
313. The authors do not advance a theoretical framework for the effect of displays or Display Bans, nor make formal hypotheses.
314. Table 1 of the study shows the dates Display Bans were introduced vary by state in Australia and by province in Canada and there were no POS marketing restrictions during this period in the UK and USA.

Appropriateness of research design

315. The ITC survey is an annual survey and not a longitudinal panel study, which places some limitations on any examination of trends over time, because of the potential variability between each annual sample of different individuals.
316. A potential strong point of the design is the existence of control groups without Display Bans (UK and USA). However, against that must be set the difficulty of comparing survey participants from different countries, albeit all English speaking, but with different media and regulatory environments and

²⁹ See: <http://www.itcproject.org/forms>.

³⁰ Henriksen, L. (2012). Comprehensive tobacco marketing restrictions: promotion, packaging, price and place. *Tobacco Control*, 21, 147-153.

possibly also different smoking behavior. With any field survey, especially over long time horizons, there is also the risk of unobserved confounding factors (e.g. price changes, other legislative changes, etc.), a topic the authors do not elaborate on.

317. One other potential confound I should mention is regulation around in-store advertising of cigarettes and tobacco purchasing. This is because the question the authors use to define exposure to POS displays actually confounds these with in-store advertising (see below). If tobacco control measures for in-store advertising were introduced at different times to Display Bans this is a potential confound (and I understand this to be the case in at least some, if not all, of these countries). Table 1 does not make this distinction.
318. The authors also acknowledge that in reality there is less useful data than might be apparent from their introductory discussion. In some Australian states, Display Bans were only introduced at the time of their last survey, meaning only initial impact could be assessed. In addition, some of their samples and subsamples are quite small, limiting statistical power, and not all the relevant measures were asked in each country and for each survey wave, limiting their ability to make comparisons on some measures.
319. Overall, because this is an annual survey, because the control groups are other countries, and because other potential confounds were apparently not considered, this is not a particularly strong design in terms of meeting the authors' objectives.

Representativeness of sample

320. The ITC survey is conducted by telephone with adult smokers (aged 18 or over, smoked at least 100 cigarettes in their lifetime and smoked at least once in the last 30 days before the survey). Random-digit dialing is used and replenishment sampling from the same sampling frame is employed to reduce non-response. However, the authors do not discuss non-response rates or characteristics, so it is difficult to assess any non-response bias that may have been introduced into these samples. They also do not discuss the issue of mobile-only households, another potential source of bias, and whether the frame includes mobiles or not. The samples obtained are approximately 2,000 respondents per wave and country, which are reasonably large samples.
321. The authors present details of their sample characteristics by country in Table 2. There are some apparent differences between the profiles (e.g. compared with other countries, the Australian and UK samples have a higher proportion of low education respondents, the USA sample a higher proportion of low income respondents, and Australia and Canada samples a higher proportions of respondents in the heavier daily cigarette consumption categories). However, no tests appear to be conducted on whether these differences are statistically significant or not.
322. The table also says their percentages are on "unweighted" data, perhaps implying that weights might have been applied to some of their subsequent analyses. However, there is no discussion of this point that I can find in the paper and therefore I am unclear what the authors are actually analyzing. As I

note at the beginning of my report, survey sample characteristics (e.g. age, gender, education) often do not resemble the target population the researcher wishes to generalize to. One solution to this problem is to “weight” the data so that it more closely approximates the target population (e.g. if we have 50 percent fewer 18 to 24 year old males than we should have in the sample, we make each of those 18 to 24 year old males we do manage to survey count for 2 instead of 1 in subsequent analysis). Weighting is a highly complex and specialized topic and has important consequences in terms of what we can and cannot say from the subsequent analyses.

323. There are two related issues here. One, if generalization to the population is the aim of the study we do not know how the reported profiles compare to the profile of smokers in each country. That is, no checks were applied on the adequacy of the sampling procedure and we do not know the magnitude of weights that need to be applied. If indeed generalization is the aim, weighting procedures might then be used to overcome any differences between sample and population. However, typically survey researchers do not like to apply large weights as that runs the risk of distorting the results.
324. Since these issues are not discussed in the paper one might assume generalization is not the aim here, though some of their conclusions seem to imply generalization is indeed the aim. Equally, if the authors did not apply weights then we are comparing country samples that are metaphorically “apples” and “oranges”.
325. In their favor, the authors do use age, sex, education, income and cigarette purchasing as control factors in their regression analyses, and this goes part of the way to resolving these problems. For that reason, I primarily focus on the regression analyses here, though it would have been preferable for the authors to have discussed these issues or been clearer on what they did with respect to these sample differences. As I discuss later, this lack of clarity is then compounded by lack of detail on the regression results.

Measurement issues

326. There are a number of issues around the measures employed here, above and beyond the problems of mentioned above. First, these are self-report measures and for some questions respondents were asked “*in the last six months*”, which is a long time to recall “*things that promote smoking*” or “*cigarettes or tobacco products being advertised on store windows or inside stores*”. Others have questioned whether asking survey respondents to recall minor events over long time periods of time yields accurate results (for example, Carter et al., 2015). Equally the type of prompted recall questions they use for various types of advertising or promotion of also been criticized as potentially leading the respondent to a desired response. Second, the actual question that was used to define participants’ exposure to POS displays is “*In the last 6 months, have you noticed cigarettes or tobacco products being advertised on store windows or inside stores where tobacco is sold?*” I would suggest that this is not measuring POS displays at all, but retail store advertising. In my opinion, the question does not define POS displays clearly enough, and also may confound the display of tobacco products with other forms of store

advertising. Others in this field also believe that any effects of, for example, in-store advertising should be identified separately from any effects of POS displays and show they do indeed have different effects (e.g. Kim et al., 2013).

327. Purchasing behavior is measured by asking whether the respondent has a regular brand and if they bought this at their last purchase. It is not clear to me what is measured if they bought some other brand. In addition, this measure does not appear to be used in the regression analyses. In Wave 6, participants are asked if, in the last six months, they ever bought a brand other than their usual brand because they noticed a POS promotion (once again, the display of tobacco products is confounded with in-store advertising). In Wave 8, smokers were asked if tobacco displays made them buy unplanned cigarettes. In my view, this question has the potential to be leading.
328. Another observation is that it is not clear what the implications of these measures are for cigarette consumption. Does a non-usual or unplanned purchase lead to greater consumption or is it simply substitution with the regular brand with no impact on consumption?
329. Finally, the authors themselves also note issues in equating education and income categories across countries, a well-known problem in international research.
330. In my opinion, the various measures in the ITC Survey are likely to be insufficiently precise on the topic of POS displays and also subject to major recall and other biases. The authors themselves acknowledge the six-month recall limitation but believe this does “*not interfere with our ability to measure the impact of POS display bans*” (p. 908). I am not totally convinced by this argument, particularly given the criticisms others in the field have leveled at such questions, and also the apparent confound between in-store advertising and POS displays.

Appropriateness of analysis and key study results

331. The authors use logistic regression models and they use generalized estimating equations for these because of their concern for correlations within participants across survey waves. Given the ITC Survey is not a panel methodology, and the chances of the same participant answering more than one wave must be low, in my opinion this last step was not strictly necessary. The need for logistic models arises because their dependent measures are binary. However, they find that these modelling approaches are only possible for Waves 7 and 8 because the samples under display bans are too small in earlier waves. This is a limitation because these models control for the sample differences noted above, whereas their other, more standard, analyses presumably do not and are less easy to interpret as a consequence.
332. **Key study results:**
 - (a) **Exposure to tobacco advertising and promotional activities.** With the above qualification, their first set of analyses basically shows stores in the four countries complied with the law. As reported, exposure to POS marketing (as they define it) goes down when and where bans

implemented, whereas this exposure remain high elsewhere. Exposure to overall marketing activity and other (non-display) marketing communications also go down in Canada, and for these other channels also in Australia. All these effects are significant at a high probability level ($p < 0.001$).

- (b) **Purchasing behavior.** They note the proportions of respondents stating they bought a non-usual brand as the result of noticing tobacco displays/advertising were low in three of the countries (around 11 percent between Waves 6 and 8), albeit somewhat higher in USA (around 24 percent in Wave 6 declining to 17 percent in Wave 8). Declines in these proportions can be seen in Australia and Canada but not in the UK. A strong point of these standard analyses is that the authors report them in detail, including confidence intervals, which is good practice.
333. **Logistic regression.** The authors devote only one paragraph to their logistic modeling analyses which, given my comments above, is unfortunate as these analyses potentially control better for country and sample differences. Their general conclusions from these analyses, which are based on Waves 7 and 8 of the survey, are that smokers in areas with a Display Ban were less likely to be exposed to “*POS tobacco displays or advertising/promotional activities in other specific channels, had a lower level of overall salience of tobacco marketing and were less likely to purchase non-usual brand of cigarettes (or buy unplanned cigarettes in Wave 8) because of exposure to tobacco displays/advertising*” (p. 907).
334. Despite this lack of detail, if one looks at their modeling results in Table 5, apart from the obvious reduction in exposure to tobacco displays following the introduction of the Display Ban, there are significant associations between the Display ban and the two outcome behaviors, bought a “non-usual brand because of displays/advertising” and “bought unplanned cigarettes due to display.” Both of these are highly significant at the $p < 0.001$ level.
335. **Effect size/behavior.** For the “non-usual” brand outcome the odds ratio is 0.35, which would be a *moderate* magnitude effect using my calculations from paragraph 42 above. For “unplanned” purchase the odds ratio is 0.39, which would also be a *moderate* magnitude effect.

Overall conclusions

336. In reality, despite the motivation in the introduction to the paper this is not, in my opinion, a strong longitudinal research design, rather the analyses presented by the authors represent a series of cross-sectional country comparisons at different points in time. As such, I would suggest that serious issues arise in comparing country samples as these have, as reported, quite different characteristics. To be fair to the authors, this is a common problem in international research and not one with an easy solution. However, it would have been useful for the authors to discuss these issues in more depth.
337. The results of the logistic modelling utilized in the study, with various control factors, go some way to overcoming these sample differences but it is not clear

whether the “purchasing behavior” dependent variables measure changes in consumption or simply substitution. Moreover, the proportions of respondents who make these non-usual or unplanned purchases are relatively small compared with the many who report no such purchases.

338. There are also issues of potentially unobserved confounding factors, common to field surveys, which are not captured by demographics, especially across the five years and four different countries studied.
339. More critically, others in the field have raised questions about the reliability and validity of the type of self-report questions used in the ITC Survey, both from the point of leading questions and lengthy recall periods. On this basis, it may have been useful for the authors to discuss these issues in more depth than they do. It may also have been useful to have a definition of POS displays and supporting measures that clearly separate POS display effects from those of retail advertising in general.
340. In conclusion, while the study has four countries, large samples and at least two waves of data that are useful for more controlled analyses, it does not, in my opinion, provide reliable evidence that a Display Ban would have any impact on actual smoking behavior, particularly smoking initiation in minors and smoking cessation in minors and adults.

Judy Li and Darren Walton. The immediate and long-term changes in support of a tobacco display ban after its implementation in New Zealand. *Tobacco Control* 24.e1 (2015): e98-9. (Li and Walton, 2015)

Objectives, background and theory

341. This study is not about changes in individual smoking behavior per se, rather it concerns support for a Display Ban amongst smokers. Such support can be argued to be part of the surrounding environment to smoking behavior and so it is useful to include one study from that stream of research. However, the study is also interesting because it uses data from an omnibus panel, a longitudinal methodology that is common in marketing research but less so in the display ban literature. The difference is that in a panel the same people are interviewed at different points in time whereas in most research fresh samples of people are obtained each time.
342. The study is also short, two pages, so cannot be expected to contain as much detail as others I have reviewed.
343. The authors state their objective was to look at how support for the Display Ban changed from immediately before, immediately after and one year after the introduction of the ban in New Zealand in July 2012.
344. Given its short length the study is primarily focused on methods, analysis and results, with only a brief mention on previous studies, which have shown support for bans to be high in Canada and low in Norway but not particularly associated with whether a ban had been implemented or not (Brown et al. 2011; Scheffels and Lavik, 2012).

Appropriateness of research design

345. The New Zealand Smoking Monitor used by the authors of this study is a fortnightly telephone survey which recruits its respondents from an omnibus panel of adults who commit to be interviewed up to six times. Freshly recruited respondents then replace those who drop out or reach their six interviews.
346. The authors state that they analyzed data from 726 respondents who completed a total of 2,160 interviews. Some of these interviews were in a two-month period immediately before the ban, some in a two-month period immediately after the ban and some interviews one year after the ban. However, in terms of interviewing the same people at multiple points in time only 86 individuals took part in interviews immediately before and after the ban. This is limitation because a more balanced omnibus design might have had a larger group of the same people interviewed at different times, together with others interviewed uniquely at one point in time as a contrast and control on any repeated measure biases.

Representativeness of sample

347. The omnibus panel is claimed to be nationally representative of New Zealand adults, but the authors oversample recent quitters to allow comparisons with smokers. On the one hand that is a logical step, on the other it complicates

our ability to generalize to the population, a point the authors mention in their conclusions. Panels also have the potential threat that the responding panelists are in some way behaviorally different to the rest of the population, that is, they are people who like to complete surveys, may have time on their hands, etc.

Measurement issues

348. The authors do not discuss their questions in any depth. One can infer from the analysis that the dependent variable support measure is probably a simple yes/no. They also include measures on recent quit attempts and socio-demographic factors as controls in their regression analyses. The recall period that respondents are requested to consider for quit attempts is three months, which is perhaps too long given criticisms made in other studies and the fact the survey is administered fortnightly might have allowed them to use shorter recall periods.
349. The authors themselves raise the issue of whether repeatedly interviewing (some) respondents over a short period of time might create response bias. They believe this risk is minimized by the statistical methods they use. While I agree, I would also have preferred a larger sample of repeat interviews contrasted with one-time only responses. If the former can be shown to be relatively unbiased then they allow stronger conclusions on change, because we are less concerned with controlling for individual differences. Equally, if there is bias observed, one might use differences-in differences methods to control for this (see Irvine and Nguyen, 2014).

Appropriateness of analysis and key study results

350. They follow similar analysis procedures to Li et al. (2013), namely looking at basic results and then (presumably) running a logistic regression with control variables. Like Li et al they also use generalized estimating equations and that is appropriate here as some respondents were interviewed more than once.
351. **Key study results.** In the raw results, support for the Display Ban did not change significantly from immediately before to immediately after the ban (64 percent versus 54 percent), but it did increase significantly one year after the ban (to 80 percent).
352. After entering the various control variables into their model, the differences between immediately before the ban and one year after the ban are still significant ($p < 0.05$). Examining their control variables they also note that those who had made a recent quit attempt or had been interviewed more often were more likely to show support. They do not discuss the latter result, which could potentially be some form of response bias as they mentioned before.
353. Their overall result differs from the Norwegian study by Scheffels et al. (no change) and Canadian studies (no differences between provinces with/without bans) and they suggest this may have something to do with the surrounding tobacco control environment, particularly in Canada. This is important as it reinforces the point that there are many factors influencing individual behavior and research designs need to consider these.

354. **Effect size/behavior.** There is no behavioral dependent variable in this study, since the focus is on support for the display ban. The odds ratio for the increase in support one year after the ban (versus immediately before the ban) is 2.2, which is an effect of medium magnitude.

Overall conclusions

355. This paper makes some contribution to the discussion on Display Bans in that it: (1) suggests there may be a lagged effect of Display Bans on support for them, and (2) notes the issue of the surrounding regulatory environment as a possible influence on support. I would extend that point to the broader media and social environment, topics other studies in this area would do well to consider and incorporate into their research designs.
356. The authors also make a useful point on the importance of an adequate sample of recent quitters. While this does complicate analysis and generalization to the population, they note recent quit attempts may predict “*responsiveness to policy and future quit attempts*” (p. e98). It is therefore important to include adequate samples of recent quitters in research of this nature.
357. That said, in my opinion, the research design did not take best advantage of the opportunities provided by having repeated surveys at short intervals. For example, one could have tracked support fortnightly with larger samples and also utilized the differences between those interviewed more than once and those interviewed only once to better understand the phenomenon.
358. Concerning the objectives of my report, my conclusions are as follows. (a) *Smoking initiation in minors.* This study focuses on adults, not minors. (b) *Smoking cessation in minors and adults.* This study does not measure smoking cessation. It is primarily useful in that it points out the importance of the surrounding environment and the potential higher sensitivity of recent quitters, suggesting additional factors that might be incorporated into future research.

Ann McNeil, Sarah Lewis, Casey Quinn, Maurice Mulcahy, Luke Clancy, Gerard Hastings and Richard Edwards. Evaluation of the removal of point-of-sale tobacco displays in Ireland. *Tobacco Control* 20.2 (2011): 137-143. (McNeil et al., 2011).

Objectives, background and theory

360. The stated objective of this study is to evaluate the short-term impacts of the Display Ban in Ireland, which was introduced in July 2009.
361. The authors review the literature, particularly Wakefield et al. (2008), Carter et al. (2009) and Paynter's review (2009), arguing this literature suggests POS displays are more a trigger to purchase than about brand choice and, as such, more likely to influence youth than adults.
362. The authors then present a "logic model" (Figure 1) for evaluating the impact of the Display Ban in Ireland. This they adapt from similar models in the literature, particularly those of Edwards et al (2008) and Haw et al (2006). This is a useful step in outlining the inputs, process indicators and outcomes (short, medium, long term) potentially arising from a Display Ban. However, I should note this model is at a high level and lacks sufficient behavioral grounding to be considered as a "theory" as such, or even as a testable model in the normal sense used by social scientists. For example, the whole of the discipline of consumer behavior research is contained in one box of this logic model ("*changes in knowledge, attitudes, etc.*"). That said, in the outcomes part of Figure 1, the authors advance some informal hypotheses and state over what time horizon they should emerge, which is also a useful step.

Appropriateness of research design

363. The study includes audits of retailer compliance and surveys of both adults and youth. The overall design is a simple before and after study, with the Display Ban as a single event in time. This is a limitation compared with studies in Australia or Canada, where the federal and province structures provide multiple events at different times, but it is obviously one forced on the authors by the Irish situation. The problem is that this situation does not readily allow a control group (individuals not exposed to the Display Ban) since all retailers in Ireland had to comply as of the same date, and this adds considerably to the difficulty of controlling for unobserved factors in any analyses.
364. The adult data comes from monthly omnibus telephone surveys conducted by a market research firm.
365. For smoking prevalence amongst adults, the authors have survey data for 84 months before the ban and 12 months after, which they state allows them to analyze these data as an ARIMA interrupted time series. However, to me it is not clear how this part of the study or ARIMA analysis was conducted as the authors do not discuss it in detail and only mention the results in one sentence (page 140). For other dependent measures, it appears the authors combined the monthly surveys before and after the Display Ban into two data sets and performed simpler analyses (page 138). The total sample sizes for these

analyses are not clear to me. Indeed, in general it might have been useful if the editors and reviewers of the journal had insisted on more details of these aspects of the study.

366. For the youth population, they have one survey administered in June 2009 by the same market research firm, and then in August 2009 they re-interview the respondents from the June survey. So here they obtain two measurements, one month before and one month after the Display Ban.
367. The authors mention that their results may be confounded by other tobacco control measures and mention the banning of small pack sizes and an increase in taxes during the study period. They do not appear to consider these as important, whereas others might regard the combined effect of these in raising minimum expenditure and prices as important factors to control for in any analyses (e.g. Irvine and Nguyen, 2014). The authors do not mention any other potential confounds.

Representativeness of sample

368. The monthly omnibus telephone survey used in this study recruits 1,000 respondents via random digit dialing with stratified quotas on gender, age, social class and regions. The fact that quotas are used means this is not a random survey, a limitation the authors acknowledge. On the other hand it allows the survey results to be generalized to the overall population through weighting procedures, which the authors do. However, they do not discuss the statistical implications of weighting or whether the analyses they present are un-weighted or weighted.
369. There is also no discussion of non-response (mobile only homes, people not at home when called or refusals to participate in the survey), so it not possible to assess whether there is any bias caused by these and/or whether this is reduced or exacerbated by the quota weighting procedures.
370. The youth survey is recruited in a similar way but with parental permission obtained. Again there is no discussion of non-response. In addition, since these respondents are re-interviewed (with 80 percent success) this study is a panel study, whereas the adult survey presumably represents a fresh sample of respondents each month. In my opinion, it would have been useful to discuss the pros and cons and implications of the two different methodologies. The youth survey has a small sample size: 214 respondents falling to 180 in the follow-up survey. The authors acknowledge this in their discussion of limitations, regarding any findings from their youth survey as “preliminary” pending larger samples in future follow-up studies.

Measurement issues

371. The authors provide their questions in Table 1 of the paper, allowing assessment of their nature and any potential biases. In that respect, it seems likely to me that respondents knew the purpose of the survey and this may have led to social desirability biases. On the other hand, it is also possible these biases may have been of similar magnitude before and after the Display Ban, which would limit their effect on any analysis of change.

372. The authors also note that while their questions are specifically focused on POS displays, the Display Ban legislation itself also banned point-of-sale advertising and it is possible some respondents could not distinguish between the two in their responses.
373. However, a relatively strong point of these questions is that respondents are asked to recall tobacco displays they saw in the last month, limiting recall biases. They do not discuss their measure of smoking prevalence for adults, while for children it is “*do you smoke at all nowadays?*” (p. 139). In my view, this is potentially ambiguous as the young people who participated in the survey may have different interpretations of “at all nowadays”. For 10 months of the omnibus surveys, extra questions were added on recall and opinions of both displays and the Display Ban itself.
374. All the measures are clearly self-report and concern smoking. In my opinion, it might have been better to first ask some more open questions rather than directly prompting respondents on displays and Display Bans. That is, an unaided measure of the proportion of respondents noticing displays in the last month may have provided a useful addition to the questions asked and left the authors less open to criticisms of priming or leading questions raised by others in the field (Carter et al. 2015).

Appropriateness of analysis and key study results

375. The main analyses the authors present contrast proportions before and after the Display Ban was introduced, and they use chi-square tests to establish the statistical significance of any changes in these proportions. This is problematic as I assume they have large sample sizes and the chi-square test is overly sensitive to these, meaning that even small differences in proportions will be significant.³¹ Hence the statistical significance they report (which is uniformly high at $p < 0.001$) is likely less important here than the magnitude of the changes themselves (effect sizes).
376. As noted before there are insufficient details provided on the ARIMA interrupted time series models. This is an oversight as, in principle, this modeling should be more rigorous than the simple analyses they present on the basic data, since it would allow them to control for various factors that may drive these data. Nor are the weighting procedures (if any) and their possible effect on study outcomes discussed.
377. **Key study results.**
- (a) **Retail audits.** Though it is not the focus here, it is worth noting that the retail audits, also conducted by the same market research firm, show very high (97 percent) compliance with the Display Ban legislation.

³¹ The size of the statistic is “directly proportional to the size of the sample, independent of the strength of the relationship between the variables.” (p. 236) Leon-Guerro, A. & Frankfort-Nachmias, C. (2014). *Essentials of Social Statistics for a Diverse Society*. 2nd edition. Sage Publications.

- (b) **Before and after comparisons of surveys on adults.** These show a 27 percent drop in the percentage of respondents recalling seeing tobacco displays in the last month (from 49 percent before the ban to 22 percent after). The 95 percent confidence limits on this drop are 25 percent to 29 percent so this is a significant and strong raw effect. The effect of the Display Ban on recall of displays was greater on smokers than non-smokers (13 percent drop versus 3 percent). Their data also suggest that support for the ban increased shortly after its introduction and this support was sustained (at around 66 percent). The authors note “there was a small but statistically significant fall in the proportion of adult smokers who thought removal of tobacco displays had made it easier to quit smoking” (from an average of 20 percent in the three months before the ban to a mean of 14 percent in the seven months after). The authors also comment that this “*easier to quit*” statistic fell to a low of 10 percent three months after the introduction of the Display Ban.
- (c) **Adult smoking prevalence (possibly the results of ARIMA modeling).** Despite the changes in raw recall and opinion, the authors find no significant change in smoking prevalence associated with the introduction of the Display Ban. I assume here they are referring to their ARIMA interrupted time series analysis, which in principle should be the more rigorous of their analyses, but this is not clear from the paper.
- (d) **Youth surveys.** It is important to note that the results here have to be qualified, not only because of the small sample size, but also because before the introduction of the Display Ban, only 10 percent of respondents reported that they were current smokers, and only another 7 percent that they are past or infrequent smokers. After the Display Ban, these percentages become 11 percent and 8 percent and the differences before/after are not statistically significant. While one cannot place too much reliance on these “preliminary” results, they are consistent with the adult surveys in the sense that no change in behavior is observed following the introduction of the Display Ban. The results on the recall of tobacco displays are also consistent with the adult survey in that this percentage dropped from 80 percent before to 22 percent after the ban. However, the percentage of participants who thought that the removal of displays made it easier for children not to smoke did not show a statistically significant change (47 percent before, 38 percent after, $p < 0.07$).
- (e) **“De-normalization”.** Despite there being no change in behavior or opinions on whether the Display Ban makes it easier not to smoke, the proportion of participants who thought that more than 20 percent of their age smoked decreased significantly ($p < 0.001$) from 62 percent before the ban to 45 percent after. In my opinion, it would have been useful to discuss the pros and cons and meaning of this measure in more depth, as it is this result that the authors highlight in their discussion and conclusions.

378. **Effect size/behavior.** The changes in smoking prevalence before and after the Display Ban are not significant ($p < 0.41$) and so it is not appropriate to examine effect sizes.

Overall conclusions

379. The authors suggest that retailers complied with the law and as a consequence exposure to tobacco displays decreased. In many ways that is expected, though it is a good question why one in five respondents still report that they see tobacco displays after the Display Ban, a phenomenon that has been observed in other studies (for example, Li et al. 2013) and which may go to the difficulty of accurately measuring recall through simple survey questions. The authors themselves suggest it is “*of some concern*” (p. 141) as some respondents may be recalling the more distant past or confusing displays with other activities at the store.
380. The authors note that their Figure 1 “hypotheses” suggest that any changes in adult smoking prevalence should emerge over a longer time scale than they studied, which they argue is primarily because of the difficulties of quitting smoking. However, I would note that some econometric studies, albeit on a different country, have shown no such longer-term effects (e.g. Irvine and Nyugen, 2014), albeit these latter studies only investigate one to two years after the ban. As I noted in my earlier discussion, there is a need for longer-term follow-up studies on Display Bans.
381. Finally, through circumstance, the authors have only one Display Ban event to study with no control group. These are important limitations on their research design. That said, the monthly adult omnibus survey is a strong point of this study and could perhaps have been used to better effect by a more sophisticated cross-sectional time series design than the simple before/after comparisons on proportions as presented here.
382. In summary, and addressing the objectives of my report, this study does include minors but has no specific measure of smoking initiation as a concrete behavior. As noted by Irvine and Nguyen (2014), the measure they do include here, smoking prevalence, is a net measure of initiation and quitting. Moreover, in this study there is no statistical association observed between the introduction of the Display Ban and smoking prevalence amongst youth. While there are measures of perception and opinion concerning smoking cessation, there is no measure of this as a specific behavior. Again, there is no statistical association between the introduction of the Display Ban and any fall in smoking prevalence. Assuming this latter result is based on ARIMA modeling, this is the strongest (non) result from the study, because this is the more rigorous analysis procedure.

James Nonnemaker, Annice Kim, Paul Shafer, Brett Loomis, Edward Hill, John Holloway and Matthew Farrelly. Influence of point-of-sale tobacco displays and plain black and white cigarette packaging and advertisements on adults: evidence from a virtual store experimental study. *Addictive Behaviors* 56 (2016): 15.(Nonnemaker et al., 2016).

Objectives, background and theory

383. The stated objective of this study is to examine the potential impact of a Display Ban and of mandating plain packaging at the POS through an experiment. This experiment uses the same virtual convenience store approach as Kim et al (2013) and Kim et al (2014) and focuses on adult smokers and recent quitters. The main outcome measures are (1) self-reported urge to smoke and (2) tobacco purchase attempts at the virtual store checkout.
384. The authors cite a number of previous papers and studies on tobacco advertising, brand imagery and POS displays with the conclusion that these may act as cues to smoking, encourage smoking initiation or result in the relapse of recent quitters to which they add a discussion of the plain packaging literature, citing for example studies by Wakefield et al. (2013), Gallopel-Morvan et al. (2012) and Hammond et al. (2013), amongst several others. The authors suggest that plain packaging may make cigarette smoking “*less attractive, reduce the urge to smoke and make health warnings more prominent*”. They also cite their own previous experimental work (i.e. Kim et al. 2013 & Kim et al. 2014) to the effect that enclosing POS displays reduces purchase attempts in youth and adults and reduces the urge to smoke amongst adults. In my opinion, the inclusion of plain packaging and enclosed displays as experimental conditions potentially provides a strong test of tobacco control measures at the retail level. Finally, the authors discuss a number of studies that have suggested virtual reality approaches have good ecological validity both in general and for studies of smoking behavior in particular (e.g. Bauman and Sayette, 2006³²).
385. The authors do not advance a theory of how Display Bans and plain packaging may impact smoking perceptions, motivations or behavior, nor do they make formal hypotheses. Rather they point out that measuring such effects in a real world setting is difficult and, for this reason, they prefer controlled laboratory experiments. They also note that this study builds on their previous two experiments.

Appropriateness of the research design

386. The virtual store format is described more fully in Kim et al. (2013). The experiment here has a simple 2 by 2 design with POS tobacco displays either open or enclosed and in-store tobacco advertising either full color for both text and packs or black and white plain packs and text. Participants to the experiment were then randomly assigned to one of the four resulting

³² Bauman, S.B. & Sayette, M.A. (2006). Smoking cues in a virtual world provoke craving in cigarette smokers. *Psychology of Addictive Behaviors*, 20(4), 484-489

conditions, providing control and test groups for both displays and in-store advertising.

387. One important point to note about this experiment is that in-store cigarette advertising is always present at the checkout counter (Figure 1, p. 17), the difference being whether it is full color or black and white. It could be argued this is a weak contrast, especially compared to the display whether cigarette packs are either visible or hidden at the checkout.
388. As in the previous studies, participants are asked to complete a shopping task involving the purchase of one item each from the snack and cooler areas of the store and two from the checkout area. If the participant asked to purchase tobacco products, the animated virtual cashier queried whether this was their intention and if it was, a list of available brands was presented. On completing their purchase, participants are then directed to a follow-up survey.
389. As with their previous two studies, in my opinion this is a simple, well-designed experiment with a commendable attention to detail. As before it is therefore likely to have high internal validity and reasonable ecological and external validity, albeit the latter is limited by the experiment being of one store visit (see my discussion of Kim et al. 2013 above). The authors provide data from the survey on ecological validity, with participants being split but those who thought it very much resembled a real-world convenience store (51 percent) and those who thought it only a little or somewhat resembled a real store (45 percent).
390. The authors also provide breakdowns of the time spent in the shopping aisles (average of just under 4 minutes) and at the checkout (32 seconds), with only a small variation between participants. The authors note the latter is “*considerably less*” (p. 21) than the one minute people spend at the checkout in real stores and this may be one limitation on the external validity of their study. However, as in the previous two studies, the authors point out their focus is more on internal validity than external validity. That is, if an effect is observed in their experiment, one can be confident it exists, but one is less confident that we can generalize this effect to the real world.

Representativeness of the sample

391. The authors note that their sample is a convenience sample and therefore cannot be generalized to the overall population. They use similar procedures as before to recruit their sample of current smokers and recent quitters aged 18 years and older. I would also note a further limitation on generalizability is, as in the previous studies, a large drop-out rate between the 4,412 adults who agreed and were eligible to participate in the experiment and the 1,313 who eventually completed the shopping task and survey. If one were to seek to generalize, an important question to answer would therefore be whether these 1,313 are in some way different to other members of the population.
392. The authors present the profiles of the participants in each of the four experimental conditions and there were no statistically significant difference between these profiles, demonstrating that their randomization procedures were effective.

393. 64 percent of the 1,313 participants were current smokers and 36 percent recent quitters. These participants were obtained from an online panel run by a market research agency. Again if generalization were the goal, this would require some examination of potential panel biases.

Measurement issues

394. It appears the purpose of this study was well disguised in that participants would not necessarily be aware of it until after the virtual store task. Hence the purchase attempt measure is much less likely to be socially biased than the questionnaire measures. However, the authors do not provide adequate details on what participants were told prior to the study, so here I'm surmising that they followed good experimental practice to minimize any such biases.
395. The "urge to smoke" measure in the questionnaire is a self-report measure and therefore more open to biases. It is also more of a measure of *motivation* rather than a measure of behavior. In contrast, purchase attempts are a measure of a concrete behavior and a behavior that is accurately measured by the purchase confirmation and brand selection steps in the checkout task.
396. The survey includes questions on awareness of the store advertising, POS displays and health warnings and also measures a number of control factors for use in regression modelling, in particular: age, ethnicity, gender, education and frequency of going to a convenience store.

Appropriateness of analysis and key study results

397. The authors' main analyses of interest here use logistic regression techniques because the dependent purchase attempt measure is binary. Ordinary regression techniques were used for the urge to smoke measure as this was measured on a 0 to 100 scale. They estimated models for the main effects of enclosed displays and plain packaging and also for the interaction between the two. However, the latter is never significantly different from zero so the authors dropped it from their analyses. Separate models were run for current smokers and recent quitters. However, in the two enclosed display conditions recent quitters made very few purchase attempts and so the effect of the display ban could not be estimated for this group.
398. **Key study results.** Here I focus on the results for purchase attempts and only briefly mention the findings on the urge to smoke measure.
399. In terms of raw effects, significantly fewer participants clicked to purchase cigarettes when the display was enclosed (for current smokers, 3.8 percent or 2.9 percent compared with 51.4 percent and 24.4 percent in the corresponding open display conditions, $p < 0.01$, p. 20). Also within the two open display conditions, those in the plain pack and advertising condition were significantly less likely to make a purchase attempt than in the color pack and advertising condition (24.4 percent versus 51.4 percent, $p < 0.01$).
400. For recent quitters, and the two plain pack and advertising conditions, more participants in the open display condition made purchase attempts than in the enclosed display condition ($p < 0.05$).

401. There were no statistically significant differences for most of the comparisons on the urge to smoke measure, except the urge to smoke was lower in the enclosed display and plain pack condition than the open display and plain pack condition ($p < 0.05$).
402. **Logistic regressions.** I now turn to the modeling results. These models include the control factors mentioned above i.e. factors that might also correlate with the raw differences just discussed. Accounting for these in the models provides a clearer picture of the effects of enclosed displays and plain packaging. Given the small number of purchases, the model on recent quitters could only contrast the two pack and advertising conditions, finding no significant differences between them. For current smokers, the enclosed display had a significant effect on lowering purchase attempts ($p < 0.01$), as did plain packaging and advertising ($p < 0.01$). This shows that, after controlling for various factors, there is a significant overall association between both enclosing the display and plain packaging and advertising in lowering purchase attempts in a virtual store experiment.
403. The enclosed display also lowered urge to smoke amongst recent quitters in that ordinary linear regression analysis for that measure ($p < 0.01$). There were no significant effects on urge to smoke for either display or for packaging and advertising for current smokers.
404. As I did with Kim et al. (2014), I would also point out that the plain packaging and advertising conditions confound two different marketing vehicles. The authors are therefore correct in being careful not to separate the two in their conclusions. As before, it might also have been useful to present the coefficients of the covariates and, given the large sample size, to conduct tests at a higher level of significance.
405. **Effect size/behavior.** The odds ratio from the regression equation for (reducing) purchase attempts by current smokers is 0.05 for the main effect of enclosed displays. From my calculations in paragraph 42 earlier, this represents a *large* magnitude effect.

Overall conclusions

406. This study follows the approach of their previous work closely. Overall, the study shows that, in further virtual experiment, enclosing the POS display lowers purchase attempts by adults who are current smokers. This study adds to that of Kim et al. 2014 in two ways: (1) it replicates the finding on Display Bans for adult smokers; and (2) does so in the presence of plain packaging and advertising conditions as compared with the graphic health warnings of the 2014 study. This is important because it shows that, in virtual experiments, the effect of enclosing the display is separate and independent of the effects of certain other in-store factors.
407. The authors discuss various limitations to their work, notably the short exposure, whether or not participants were able to see the health warnings on the tobacco packaging, whether behavior in the virtual store resembles that in the real world, the self-report nature of the urge to smoke scale and limitations on generalizability from using a convenience sample. All of these are valid

concerns, to which I would add the drop out from being willing and eligible to participate to actually completing the survey, further limiting generalizability.

408. That said, the authors have in their three published studies to date, introduced an innovative and useful way of examining behavior in under different display and advertising conditions and with tighter controls than many typical studies on this topic. Their approach is worth refining and extending, and also seeking to improve its external validity.
409. With respect to the objectives of my report, this third study replicates and extends the results of the earlier experiment on adult subjects, demonstrating that enclosing the POS display results in a large decrease in purchase attempts. That said, this is for one visit to a simulated store and furthermore purchase attempts are neither smoking initiation nor cessation behaviors. Hence, while this is a strong effect, it is not clear to me how this relates to the ongoing purchase and consumption of cigarettes in the real world.

J Paynter, R Edwards, P.J. Schuler and I McDuff. Point of sale displays and smoking among 14-15 year olds in New Zealand: a cross sectional study. *Tobacco Control* 18.4 (2009): 268-274. (Paynter et al., 2009).

Objectives, background and theory

410. The authors state their aims are to “*examine the association between exposure to tobacco displays at the point of sale and teenage smoking and susceptibility to the uptake of smoking*” (p. 268).
411. The authors review the history of legislation restricting POS displays in various countries and cite some of the previous research on exposure to displays and smoking behavior. In particular, they note that research conducted in the USA suffers from the limitation that it is difficult to separate the effects of the POS display from the other forms of in-store tobacco advertising that are permissible under the legislation there. In contrast, for New Zealand, at the time of their study, tobacco promotion was limited solely to displays of the product at the POS, potentially allowing them to identify the effects of exposure to this display without any such confounding factors.
412. The authors provide no theory or hypotheses as such, moving directly to outline their methodology after a relatively brief introduction.

Appropriateness of research design

413. The authors invited all schools in New Zealand with Year 10 students (14-15 years old) to participate in their study. Schools that agreed were given instructions and asked to issue the survey questionnaire to all Year 10 students in class with the teacher present. Students were assured of confidentiality and of their rights to refuse to complete the questionnaire.
414. As the authors note this is a cross-sectional survey with a multi-level design, notably District Health Board, school within health board and student within school. The alignment of the data with District Health Boards would potentially allow the authors to incorporate additional information on smoking and health as control variables but this step does not appear to have been taken in this paper.
415. Instead, the authors include a number of covariates in the questionnaire itself, notably demographics and peer/parent smoking behavior.
416. Overall, this is a relatively simple research design with a survey utilizing self-report data from students and conducted at one point in time, a time at which the only tobacco promotion allowed is the POS display itself.

Representativeness of the sample

417. 47 percent of New Zealand secondary schools agreed to participate and in total 45 percent of the national Year 10 population of students completed questionnaires. As a result 27,757 completed questionnaires were obtained which is an impressive sample size, but unfortunately there is no detailed discussion or analysis in the paper of how response varies by health board, class or school, except in that the authors note: (1) the overall demographic

profile of the sample matches that of the target population; and (2) schools from areas with lower socio-economic status are under-represented in their sample.

418. Overall, the sample therefore seems to be reasonably representative of most New Zealand Year 10 students. It would, however, have been useful for the authors to do some follow-up research on the reasons for non-response, if only to rule out potential biases in the data. There are some 55 percent of students who did not complete the questionnaire and, while we know they have a reasonably similar demographic profile to those who did complete it, we do not know if they differ on other factors of relevance here. Burton et al. (2012) note that respondent self-selection into research studies can potentially lead researchers to incorrectly attribute effects to study variables, when these effects are in fact due to unobserved differences between respondents and non-respondents.

Measurement issues

419. The survey measures relating to smoking are self-reports of smoking status and smoking susceptibility. The authors do not discuss how the questionnaire is introduced and/or their measures defined to the students but they do present their questions. For smoking status these directly raise smoking as the focal topic (“have you ever smoked (even with just a few puffs)” and “how often do you smoke now”). These are then used to define students as “never smokers”, “experimental smokers” (those who had smoked in the past or who smoke infrequently) and “current smokers”. Smoking susceptibility was measured by intentions to smoke in the next year, using two questions from prior research.
420. The authors then use these questions to define their outcomes of interest, namely impacts of POS displays on students who have: (1) never smoked but are susceptible to smoking; (2) those who are experimenters; and (3) those who are current smoker.
421. I would raise one issue with the “experimenter” designation, which possibly confuses those who have ceased to smoke with those who smoke infrequently (in this study less than once a month). In my opinion, these are two different behavioral states. It is also clear that susceptibility is an intentions measure and as such likely to be only weakly correlated with any eventual behavior.
422. However, the bigger issue to me is that there is no discussion of the reliability and validity of these measures or of any potential biases in them. The latter is a significant omission, particularly given the sensitivity of the topic and the presence of the teacher in the room when the questionnaires were completed. I would have preferred to see greater discussion of these issues in the paper, since the self-reports of adolescents in the presence of adults in authority positions may potentially be subject to social desirability biases, in addition to the more obvious recall biases on behavior.
423. Exposure to POS displays was estimated through two measures, store visit frequency (for various types of stores selling tobacco products) and how often they noticed tobacco when visiting these stores. For the latter, the authors

then focus their analyses on the type of store the individual student visits most often. For many countries I would argue that “noticing tobacco” is a poorly defined measure with many potential confounds from the store environment but here I agree that in New Zealand at the time of this study the only place tobacco could be “noticed” is at the display. Therefore this is a self-report measure of exposure to POS displays. Whether it is a valid and reliable measure the authors do not discuss or present any evidence on, in common with much of the research literature I review here.

Appropriateness of analysis and key study results

424. The authors use three-level logistic regression models to “*investigate the association*” between POS exposure and their three outcomes. The three-level models are appropriate given the structure of the sample, however it is not clear the various effects are incorporated into the model in the best way. For example, the large sample size would permit separate dummy variables for each health district and school, which would potentially control for more variance than their simple school socioeconomic status variable. The point at issue here being the more potential confounds you can eliminate the better your estimate of the effects of interest. Indeed, given the sample size alternative models might also have been tried, to test the robustness of estimates to other model and variable formulations.
425. The authors estimate these models with and without their control variables. Since the latter—“adjusted”—analyses better control for confounding factors I primarily focus on those for my review.
426. The sample is very large; so it is not clear the statistical tests the authors present, while necessary, add much value. With a sample of over 27,000 it is not surprising that many differences are significant at a probability level of 0.001. For this reason, I focus here mainly on effect sizes.
427. The authors also test interactions between store visit frequency and noticing tobacco in their models and between these two exposure measures and the socioeconomic status of the school. The former are not statistically significant, whereas the latter are.
428. Effectively the “control groups” for their comparisons of odds are those with low visit frequencies or those who do not notice tobacco when visiting a store.
429. **Key study results.** By way of background the authors classify 41 percent of their sample as “non-susceptible and never smoked,” 16 percent as “susceptible and never smoked,” 30 percent as “experimented” and 13 percent as “current smokers.” Close to two-thirds (64 percent) of these students visit stores two times per week or more and the majority (52 percent) notice tobacco most or every time they visit.
430. In general, the models including control variables show that those students who visit stores or notice tobacco more frequently are more likely to be susceptible to smoking, to have been experimenters or to be current smokers, compared with those who visit the less frequently or never notice cigarettes. All these effects are statistically significant.

431. **Effect size.** Here I follow the authors and use the top two categories on the exposure measure compared to the bottom reference or “control group” category. The odds ratios for the various effects in these models are then in the range of 1.8 to 2.2 for the “susceptible to initiation” outcome, in the range of 1.3 to 2.7 for “experimented,” and in the range of 2.2 to 3.5 for “current smokers.” These are mostly effects of small to medium magnitude, with only the effect on “current smokers” of those who notice cigarettes “every time” they visit a store approaching a large magnitude.

Overall conclusions

432. This study focuses on minors and has a large and probably representative sample of students aged 14 to 15 years. It therefore can potentially provide evidence against one objective of this report, namely whether it shows any evidence that a Display Ban would reduce smoking initiation among minors.
433. First, I would note that this study does not provide any direct evidence on this objective because there is no before and after measurement of the effect of any ban. Rather the study shows an association between exposure to existing POS displays and the three outcome measures. However, while the authors put forward a dose-response argument, with the implicit assumption that POS exposure (the dose) causes these outcomes (the response), in my opinion, their study cannot assign any causality to this association because it is conducted at one point in time. Indeed, in their closing paragraph they explicitly recognize this point.
434. Second, I would also note that the outcome measure most closely related to smoking initiation, “*susceptibility to initiation*”, is a measure of intention and not behavior. As discussed in paragraph 9 above statements of intention are only, at best, weakly related to future behavior.
435. Finally, the authors do not provide any evidence of the validity or reliability of their measures, which being self-reports are subject to various recall and social desirability biases. Also, in my opinion, the “experimented” measure confuses different behavioral states and also does not allow the authors to draw any conclusions on smoking cessation.
436. Overall, I conclude this study does not provide any evidence that a Display Ban would reduce smoking initiation or increase smoking cessation in minors.

Casey Quinn, Sarah Lewis, Richard Edwards and Ann McNeil. Economic evaluation of the removal of tobacco promotional displays in Ireland. *Tobacco Control* 20.2 (2011): 151-155 (Quinn et al., 2011).

Objectives, background and theory

437. The authors state their aim is to “*evaluate the short-term economic impact*” of the Display Ban in Ireland, which occurred in July 2009. The authors propose to use the available data on cigarette pack sales from retailers in Ireland to look at the impact of the Display Ban.
438. There is no theoretical framework proposed or hypotheses advanced in this study. The authors briefly mention prior research that has “demonstrated” how POS displays influence susceptibility to smoking and smoking uptake among youth (citing principally Paynter (2009) and Henriksen et al. (2010)). However, their primary focus is on an empirical analysis of sales data.
439. It is not entirely clear to me what “short-term” means in the context of this study. The authors mention “*up to 12 months after implementation*” but their data appears to end in 2009 (Table 1), implying a maximum of 6 months. However, if I calculate from their number of weeks of data (232, p.152), I would conclude their analysis mostly likely runs to the end of June 2010 i.e. 12 months after the introduction of the Display Ban.

Appropriateness of research design

440. The research design is simple in that the authors obtain time series data on retail cigarette sales from two sources, (1) weekly sales electronically scanned at checkouts in major retailers, and (2) for smaller retailers bimonthly audits of purchase records. Both these are provided by a major market research company, AC Nielsen. The authors then treat the Display Ban as a discrete event in the time series and seek to understand its impact on sales. The time period they examine is from January 2006 for the scanner data and from January 2007 for the retail audit data, and for both to the end of 2009 (Table 1).
441. As the authors note, the retail audit data, being bimonthly, has fewer data points and cannot be subjected to rigorous analysis, only visual inspection. I would concur with that, also as I note in “Measurement Issues” below there are problems in using retail audit data as an indicator of short-term impacts.
442. In general, I would suggest that this is not a strong research design, as the Display Ban was introduced nationally at the same time and there is thus only one “event”, and equally no possibility to use regional control groups. This places limitations on the analysis because of potentially unobserved confounding factors. The authors are aware of this, and in their limitations section mention several possible confounds, including other tobacco control measures, increased taxes and counterfeit cigarettes which may have affected demand over the period.
443. What I find puzzling is that the authors do not include the price of cigarettes in their design. The price the customer paid is readily available in electronic

scanner data and would have been an important factor to include, controlling for price changes and/or tax increases over the period of analysis.

Representativeness of sample

444. In the context of this study, the sample is one of the sales of cigarettes in retail stores of various formats. As the authors note, cigarettes are sold in a variety of retail formats, including most prominently “multiples, “symbol groups” and “forecourts,” but also in smaller outlets such as newsagents. Their scanner data covers the larger formats, whereas the audit data covers the smaller ones. They note their data does not include vending machines, and one major retailer chain is omitted for reasons that are not stated.
445. However, as analyses they present show, the authors have reasonable coverage of the majority of cigarette sales made in Ireland. In particular, the weekly electronic scanner data covers the majority of cigarette sales, and this is a strong point of the study because of the accuracy of these data.

Measurement issues

446. In one sense, electronic scanner data provide highly accurate records of actual behavior: we know someone bought a specific product for a certain price at a known location, date and time. However, how useful these data are for studying individual behavior depends on the nature of the product and what is done with these records afterwards. I note this paper analyses *sales data*. Sales data are typically aggregates of the individual electronic records from checkouts and retail stores. They are prepared for managerial purposes and cannot easily be equated to the behavior of individual customers. This is particularly the case with frequently purchased products like cigarettes where differing pack sizes, multiple purchases made by a customer at the same time, and purchases made by customers who visit the store more or less frequently, have all been aggregated into one overall weekly sales number. If one wanted to study individual behavior it would be better to use *scanner panel data*--where a customer identification code is entered at the same time as the purchase data (e.g. from a store loyalty card)—connecting the individual to the amount and frequency of their purchases.
447. Aggregation is also a problem with the retail audit data, which is combined into 8 or 10 week blocks, further disguising the connection between individual behavior and the data provided (e.g. the data might include customers who buy every week and customers who buy every month). Aggregated data like these also smooth out the short-term fluctuations in demand that are of real interest here given the objectives of the authors. Because of this limitation, in sophisticated market modeling, retail audit data has largely been abandoned in favor of scanner sales data and scanner panel data. Hence I will focus on the authors’ analyses of weekly scanner sales data.
448. I mention these points because, while it is legitimate for the authors to claim that aggregate sales have or have not been impacted by a Display Ban, making the link to the specifics of individual behavior is more difficult and, in my opinion, not justifiable.

Appropriateness of analysis and key study results

449. The authors use the first differences of the weekly scanner data to remove the overall trend in cigarette sales in Ireland and they include seasonal and holiday effects in their time-series regression models (since sales might be affected by these).
450. After first differencing, appropriate statistical methods were used to check the trend had been removed and also to look for any lagged effects. The Display Ban itself was modeled as a simple binary effect. The authors' analyses were repeated for sales in the various retail formats as well as the overall sales data for Ireland.
451. The binary effect for the Display Ban is the simplest representation but given the accurate weekly picture of sales the authors had and the multiple retail formats it would have been interesting to (1) try other formulations of the event effect (e.g. the immediate, continuing and one year delayed impacts of the ban) and (2) test the effects of the different retail formats in an overall model of sales, which would have greater statistical power than the separate analyses of these formats.
452. It also appears the data from individual stores was aggregated into overall time series. (I conclude this from their statement that they run their regressions with 232 data points, which equals the number of weeks, but not the number of stores, which is far greater). Again, it would perhaps have been more interesting to look at data from individual stores.
453. In general, the authors do not provide much by way of details on all these analyses and tests, just a table of the regression coefficients for the Display Ban effect.
454. **Key results.** The authors find *no statistically significant effect* of the Display Ban on cigarettes sales, either overall or within the separate retail formats of "multiples", "symbols" and "forecourts" (covering the majority of cigarette sales in Ireland). This leads them to conclude that there is no short-term impact of the ban in the 12 months after its implementation.
455. **Effect size/behavior.** Since the effect on cigarette sales is not statistically significant it is not appropriate to consider the effect size.

Overall conclusions

456. This study is primarily interesting because of its use of scanner sales data. Other econometric studies using time-series analysis methods have typically used self-reports of smoking prevalence or consumption derived from surveys (e.g. Irvine and Nguyen, 2014). Thus, this study provides a good complement to those studies and reaches a similar conclusion, namely that there is no observed correlation between the introduction of a Display Ban and any change in the sales of cigarettes, at least in the short-term.
457. I should note that, despite this apparent lack of impact at the aggregate level, one is not able to reach any strong conclusions about individual behavior. While it might be thought unlikely that the Display Ban increased the

consumption of some individuals, while decreasing the consumption of others, we cannot rule that out as an explanation when we only have aggregate sales data. More critically, the authors do not control for price, which they could have done with these data, at least to rule out price changes as a factor potentially masking the impact of the Display Ban. I make these points because the authors, in their conclusions, return to individual behavior research to argue that despite this lack of immediate impact, the Display Ban is “likely” to have a longer-term impact on reduced sales and smoking initiation among youth. To me, this is pure speculation and should not be taken as based on evidence derived from their data or analysis.

458. A final point is that I do not believe the authors took best advantage of their data. In addition, to the point on price made above, I would also suggest that analyses at the individual store level would have been a better direction to pursue. Individual store data can be linked to catchment area statistics and allow more nuanced analyses and inferences, as well as providing greater statistical power. It would, of course, be better still to obtain scanner panel data from this period which, coupled with the right analytical techniques, would yield more precise conclusions on individual behavior as opposed to aggregate sales.
459. In summary, I conclude this study does not tell us anything about reducing smoking initiation in minors (as the smoking behavior of minors is not identifiable in these aggregate sales data). Further, this study provides only circumstantial evidence about smoking cessation in minors and adults. Here, what we can say is that the authors observe no statistically significant correlation between the introduction of the Display Ban and the sales of cigarettes. While this is indeed suggestive of no change in smoking cessation rates, we lack the detailed picture of individual initiation, consumption and cessation that lies behind these aggregate sales numbers and which might allow us to make stronger conclusions.

Janne Scheffels and Randi Lavik. Out of sight, out of mind? Removal of point-of-sale tobacco displays in Norway. *Tobacco Control* 22.e1 (2013): e37-42 (Scheffels et al, 2013).

Objectives, background and theory

460. The authors state their objective is to evaluate retailer compliance and consumer perceptions and experience with the POS Display Ban introduced in Norway in January 2010. Given the objectives of my report I will focus more on the consumer perceptions and experience component of their study and only briefly discuss retailer compliance.
461. The authors suggest that POS displays are an important media channel for tobacco manufacturers and they cite some of the literature on POS displays, including the review of Paynter (2009) and studies conducted in Australia (e.g. Wakefield et al, 2008; Carter et al, 2009).
462. They also note that Norway has implemented tobacco control measures from the 1970s onwards and the Display Ban introduced in 2010 is a continuation of these measures, with the specific aim of preventing youth smoking initiation and supporting cessation attempts.
463. There is no theoretical framework proposed or hypotheses advanced in this study. Indeed, the introduction is short (one page) and the authors then move directly to their methods.

Appropriateness of research design

464. The authors use three methods to collect their data: a post-ban retail audit, web-based surveys and focus groups. The retail audit is of a sample of grocery shops carried out by Nielsen Norway and the three web-based surveys of consumers were conducted by another market research agency, TNS Gallup. Both use questions designed by the authors. One survey was conducted before the ban, one immediately after the ban and one 10 months later. Five focus groups were conducted before the introduction of the Display Ban and five after.
465. In short, this is a simple before/after ban design and as the Display Ban was introduced across Norway on the same day the authors are not able to set up control groups. The authors do not discuss the strengths and weaknesses of this design, nor do they discuss any potential unobserved confounding factors in the Norwegian environment during 2010. Indeed, there is no discussion of potential limitations in this paper despite the fact that simple before/after designs without control groups are, in my opinion, a relatively weak approach to identifying the effects of events such as a Display Ban.

Representativeness of sample

466. The retail audit drew on a random sample of 351 grocery stores visited by professional market research observers. This represents a reasonably large sample of such stores in Norway.

467. The web survey design itself includes a sample drawn from a large commercial Internet panel and additional samples from this panel to obtain sufficient numbers of tobacco users for statistical analysis. The former having over 900 respondents for each survey and the latter over 300, again for each of the three surveys. There is not much discussion of how these additional samples are drawn but possibly they are not at random, which means that care is needed both in combining the two types of data and in drawing inferences from them. It is also not clear if any of these respondents are re-interviewed from survey to survey which, depending on their proportion in the data, might require corrections to statistical analyses or separate analyses.
468. The authors clearly state that they weight the random survey data to allow generalizations to the population of Norwegians aged 15 to 54 years. Here they also note they do not weight the additional samples since they lack data on the total population of tobacco users.
469. In this context, “tobacco users” include both smokers and snus users, which are broken out separately in the authors’ analyses. I will focus on smokers since I understand the sale of snus is largely illegal outside Scandinavia.
470. There is no discussion of response rates or tests to examine non-response bias. Respondents are also drawn from an established commercial panel and given incentives for responding, which implies some potential for panel biases. Again there is no discussion or tests of these biases. While I accept that these surveys were conducted by reputable market research agencies, if generalization to a total population is an important objective of the authors’ work then it is necessary to assess sample biases like these.
471. As discussed above, focus groups are only appropriate to generate or test ideas or add descriptive detail to studies. They do not produce reliable or generalizable evidence on effects.

Measurement issues

472. The authors detail their web survey questions in Table 1. Though I am not familiar with the valence of the word “tempted” in Norwegian, it is possible the first question in the table is a leading one in that it immediately focuses respondents on POS displays (“Do you ever get tempted to buy tobacco or snus when you see the products at the point of sales?”). This question is also only asked of smokers/snus users. Following the advice of others in this field of research (Carter et al., 2015) it may be better to ask an open question on potential purchasing influences before moving to more directive questions on displays and display bans.
473. The authors appear to use this single item as a measure of impulse purchasing. However, with only three response categories (“Yes, often”, “Yes, sometimes” and “No, never”), this is a relatively crude and potentially unreliable measure of self-reported behavior. One can also argue that “tempted” is not behavior in itself but a precursor to subsequent behavior. There is no discussion of such validity issues in the paper.

474. However, the more serious issue is that the impulse-purchasing question is only asked one time before the introduction of the Display Ban and only to the additional sample of tobacco users. Thus we have no measure of whether “temptation” decreased after the introduction of the Display Ban. In my opinion, it might have been better to have the respondent select the influences on their purchasing behavior from a neutrally-stated list and then track how these selections changed (if at all) after the Display Ban.
475. The questions on the Display Ban themselves are potentially better measures in that they define the ban clearly and appear to be (as stated in English) more neutral in tone. Furthermore, they are asked to all respondents with only minor and appropriate changes in words before and after the ban. There are three of these questions, one asking whether the respondent is “in favor or against the ban” and the other two asking whether the ban will make it easier to quit or more difficult to start smoking. However, I note these are not measures of behavior but of opinion and perception, particularly as the questions do not define “easier” or “more difficult” for *whom*; the respondent or other people?

Appropriateness of analysis and key study results

476. The proposed analysis here is straightforward, with simple chi-square tests on changes in response before and after the Display Ban, and also for comparisons of age groups.
477. The fact that most questions have few response categories might limit the application of more sophisticated techniques, though it could have been interesting to use a technique like logistic regression to control for age, gender and smoking status as many others in the field have done.
478. **Key results.** The retail audit showed high levels of compliance (92 percent) with the Display Ban legislation by the end of January 2010 (i.e. within one month). 26 percent of tobacco users reported being tempted by displays before the ban but as noted above there is no comparable measure after the ban. The 15-24 year old tobacco users reported this more often than the older age groups ($p < 0.001$). 20 percent of tobacco users felt the ban had made it more difficult to buy tobacco products and 32 percent thought it more difficult to choose a brand. However, the only age difference here was between the youngest and oldest age groups and at a much lower level of significance ($p < 0.035$). The authors note that support for the ban itself was related to smoking status, with non-smokers showing most support (70 percent) and daily smokers least (30 percent). No statistical test is mentioned for this result, though it appears likely this would be a significant result.
479. I note the authors do not really discuss their tests of changes in response before and after the ban in the text, instead they present their results survey-by-survey and contrast these verbally. However, the results and tests they present in Table 3 (right hand column p values) rather suggest that there are no statistically significant differences across the three surveys for nearly all their contrasts. Excluding snus, the only contrast of note here is that occasional smokers appear to be less likely to think the ban will make it easier to quit smoking *after* the ban than before. This effect is significant ($p < 0.004$), with

the proportion of occasional smokers expecting the Display Ban will make it easier to quit declining from 55 percent before the ban to 39 percent ten months after. This result does not appear to be discussed in the text.

480. **Effect size/behavior.** There is no behavior as such measured in this study. Furthermore, most of the tests are not statistically significant making it inappropriate to assess the effect sizes on the various opinion measures.

Overall conclusions

481. I would suggest that the design of this study is inherently limited by the introduction of the ban nationwide at one time, which prevented the use of control groups. Compounding this problem, the authors also do not measure any control variables as an alternative way of addressing potential confounds. Nor do they discuss such confounds or any other limitations on their study.
482. Another limitation is the measures themselves. Though the authors suggest in their last sentence that a Display Ban “*can contribute to reduce tobacco use*” (my emphasis added), I believe they did not measure behavior after the Display Ban, rather simply self-reported perception and opinion. In my opinion, even their measure before the Display Ban is suspect, being potentially a leading question and better thought of as a measure of a precursor to behavior rather than behavior itself.
483. The approach taken to analysis also limits the conclusions that can be made since the authors focus mostly on cross-sectional age group comparisons rather than changes over time. Indeed their Table 3 statistics show almost no statistically significant differences between the three surveys across a range of questions and contrasts, suggesting little impact of the ban itself amongst tobacco users.
484. I suggest this study does not provide evidence about reducing smoking initiation in minors, since their measures of temptation at the POS display, purchase difficulty and brand choice are for individuals who already smoke. Moreover, the temptation measure is only made *before* the ban, which does not allow us to make any conclusions about the effect of the ban on this measure. In relation to smoking cessation in minors and adults, the questions on the difficulty of purchase and brand choice might potentially provide some circumstantial evidence but the results suggest little change in these responses either immediately or 10 months after the introduction of the Display Ban. Indeed, the only significant result goes against the thrust of the paper, in that the proportion of occasional smokers who perceive the ban will make it easier to quit appears to fall after the Display Ban.

Dionysis Spanopoulos, John Britton, AnnMcNeil, Elena Ratschen and Lisa Szatkowski. Tobacco display and brand communication at the point of sale: implications for adolescent smoking behavior. *Tobacco Control* 23.1 (2014): 64-69. (Spanopoulos et al., 2014).

Objectives, background and theory

485. This study was conducted in 2011, prior to the introduction of a Display Ban in the UK in 2012 (large shops) and 2015 (small retailers). The authors state their aim is to “*investigate the association between tobacco displays and brand communication at the POS and adolescent smoking behavior, and to assess the potential benefits likely to accrue from this legislation*” (p. 64).
486. After discussing the background to the Display Ban legislation, the authors cite only two prior research studies before moving to their methodology section. These are the studies of Paynter et al. (2009) and Mackintosh et al. (2012), which the authors conclude show an association between POS Displays and smoking susceptibility and smoking behavior in adolescents.
487. As such, there is no theory or formal hypotheses advanced by the authors, and minimal discussion of the prior literature, either empirical or methodological.
488. I also note this study is based on the first wave survey of the two waves reported in Bogdanovica et al. (2015), conducted in March 2011 and March 2012. However, it warrants a separate review because it reports data from all those 5,376 students who completed the first wave questionnaire, whereas Bogdanovica et al. (2015) examines only the 2,270 students who completed both waves.

Appropriateness of research design

489. The research design employs a cross-sectional survey of secondary schools in Nottingham involving 11 schools that agreed to participate in the study.
490. Parental consent was sought and then a questionnaire administered to the children of those agreeing, for students in school years 7 through 10. Potentially this is a clustered sample design since these students are presumably members of various classes within years, though this point is not discussed here. It is discussed in Bogdanovica et al. (2015), who adjust their analyses to account for this clustering, a step that is not taken in this study.
491. This is a simple one-time survey design and therefore cannot provide evidence on causality, only on the association between POS exposure and smoking behavior.
492. The survey takes place approximately one year before the large shop Display Ban and so cannot provide any direct evidence on the impact of this ban.

Representativeness of the sample

493. There is no discussion of the implications of only 11 out of the 36 schools in the Nottingham area agreeing to participate and whether this places any limitations on the generalizability of the study. Instead, the authors use the

response rate within these schools (74 percent) to claim generalizability to the “*wider UK population*” (p. 68).

494. In my opinion, their effective response rate is 61 percent not 74 percent because 1,109 of the 6,485 who provided questionnaires did not complete the outcome and POS exposure questions. While 61 percent is still a good response rate, it is of concern that potentially this represents data from only 19 percent of the school population of interest (11/36 multiplied by 61 percent). In my view, it would have been useful for the authors to provide more data and analysis here, especially given their claim of generalizability.
495. Equally, the authors do not provide any information of any potential differences between children whose parents agreed to their participation in the survey and those who did not.
496. They do provide profiles of their sample for those who completed all questions and those who omitted to answer the key questions and which they therefore exclude from their analyses. These profiles are useful for assessing any remaining biases and those excluded “*were more likely to be male, to report average or below average academic performance and to report that smoking is not allowed in their home...*” (p. 65).
497. Overall, the authors’ claim of generalizability sits uneasily with the fact that only one-third of the schools responded, the students who did respond are different to those who did not, and many questionnaires have missing data, particularly on key questions. In my opinion, it is possible the final sample is not even representative of the Nottingham school population, let alone the “*wider UK population*”.

Measurement issues

498. The survey measures relating to smoking are self-reports of smoking among family and friends, smoking status and smoking susceptibility. The authors also included measures of self-perceived academic performance, rebelliousness, gender and age and via the student’s postcode were also able to associate an area-level measure of socioeconomic status to each student.
499. The measures of exposure to POS displays were also self-reports: of how often students go to supermarkets and small shops, how often they noticed tobacco displays and which popular brands they recognize. I note the authors do not provide any detail on how “tobacco displays” were defined to the respondents, so the accuracy with which effects are ascribed to actual POS displays is hard to evaluate. Furthermore, there is a potential for these questions to be leading ones. The actual question is not included in the study, but the text indicates it was a direct question focusing on cigarette displays rather than indirectly about what the respondent noticed in the retail outlet. Carter et al. (2015) have pointed out the potential biases that might arise from such direct and leading questions and indeed the authors themselves recognize the possible bias in self-reported recall of displays and in prompting responses to the brand recognition question (p. 68).

500. The two measures of possible interest to my review are those of *smoking status*, a self-reported behavior and *self-reported smoking susceptibility*, a potential precursor to behavior, but not a behavior in and of itself.
501. For *smoking status* the authors used a question derived from another national survey which classifies respondents into “never smokers” and “ever smokers”, where the latter category aggregates those who have tried smoking in the past with those who smoke currently. In my opinion, this is an unhelpful aggregation, since it confuses two quite different behavioral states and therefore makes it harder to establish associations with exposure to POS displays, which could potentially have different impacts on people in the two different states.
502. The questions on *smoking susceptibility* are from a previously validated scale³³ built from three questions leading to a binary classification into susceptible or non-susceptible. While the use of three questions to determine susceptibility potentially provides a strong measure, there is clearly an issue of bias from socially desirable responses, since all three questions focus on cigarette smoking.
503. The authors report that they use imputation to replace missing values in their “confounding variables”. This is a legitimate and desirable step but they provide no details on the adequacy of their imputation model, so it is difficult to assess whether this model is appropriate and robust. It is also hard to assess the extent of the missing values problem since, while the authors do provide proportions for each variable, they do not report the proportion of complete cases, which is the standard measure for assessing missing values.
504. Overall, measurement is, in my view, a significant weakness with this study, as there are the usual limitations of self-report measures, coupled with lack of clarity on definitions and questions that are potentially leading or could provoke socially desirable responses. The aggregation of distinctly different behaviors into one smoking status measure is also a weakness of the study—it would be far better to separate current smokers from those who had tried smoking in the past, particularly as, given the focus of the study on the future Display Ban, the authors are presumably mostly interested in impacts on current smokers.

Appropriateness of analysis and key study results

505. The authors provide a number of univariate analyses before conducting two multivariate logistic regressions, one on “ever-smoking” and one on “smoking susceptibility”. I will focus on the latter here, since these multivariate regressions take account of the confounding variables and thus allow a better assessment of any association between POS exposure and these two measures.

³³ Pierce J.P., Choi W.S., Gilpin E.A., Farkas A.J., Merritt R.K. Validation of susceptibility as a predictor of which adolescents take up smoking in the United States. *Health Psychology* 1996; 15: 355–61.

506. **Key study results.** In terms of raw data, some 18 percent of students were “ever-smokers” and of the “never smokers” some 27 percent were classified as susceptible to smoking. Over 90 percent of students reported visiting a supermarket or shop every week and over 25 percent visited small shops almost every day. Over 75 percent noticed POS displays “most times” or “every time” they visited a store. From the multivariate logistic regressions, ever-smoking is associated with frequency of store visits ($p < 0.001$) and brand-recognition ($p < 0.001$). Ever-smoking is not associated with frequency of noticing POS displays ($p < 0.91$ non-significant). Susceptibility to smoking is associated with frequency of store visits, frequency of noticing POS displays and brand-recognition ($p < 0.001$).
507. **Effect size.** For ever-smoking and frequency of store visits, the highest odds ratio is 2.2 (for those who visit “almost every day”). This would be an effect of moderate magnitude. For those who visit less frequently, the odds ratios are lower and these effects would be of small magnitudes. The odds ratio for brand recognition (1.05) also implies an effect of small magnitude. For susceptibility to smoking the effects of frequency of store visits and brand recognition are also small. For frequency of noticing POS displays and smoking susceptibility, the odds ratios range from 2.0 to 3.2. These would be effects of a *moderate* magnitude.

Overall conclusions

508. The large sample of school students, validated questions and broad set of control variables potentially make this a useful study, albeit the one-time nature of the survey limits any findings to ones of association rather than causality.
509. However, it is not clear the results can be generalized beyond the sample because of potential response biases. Moreover, the self-report measures and the possible errors and biases from leading questions, and/or questions that might encourage socially desirable responses, place additional limitations on the robustness of the results obtained.
510. In my, opinion the single measure of self-reported behavior, ever-smoking, confuses individuals who currently smoke with those who have ceased to smoke. More importantly, it does not provide separate measures of smoking initiation or cessation that are the focus of my report. The study also shows no significant association between ever-smoking and noticing POS displays.
511. The other measure, susceptibility to smoking, is not behavior. It is an intention. Therefore, while the study demonstrates an association of *moderate magnitude* between noticing POS displays and smoking susceptibility, it does not provide any evidence on whether this intention is related to the smoking behavior of adolescents in general, and smoking initiation or cessation in particular.
512. Finally, from the perspective of my objectives, the major limitation of this study is that it occurs before the Display Ban is implemented. Thus it cannot provide any evidence on what the impact of a Display Ban will be on either

smoking susceptibility or smoking behavior, since the ban has not yet happened.

ANNEX TWO - ADDITIONAL MATERIALS CONSIDERED IN PREPARING THIS REPORT

I set out below the additional materials I considered in preparing this report, but which do not meet the criteria I have set out in paragraph 9 to warrant detailed analysis in this report.

- ASH. (2013). *Tobacco displays at the point of sale*. In A. o. S. a. Health (Ed.).
- Basham, P. (2010). *Canada's ruinous tobacco display ban: Economic and public health lessons*. Institute of Economic Affairs. London.
- Basham, P., & Luik, J. (2011). *Tobacco display bans: A global failure*. *Economic Affairs*, 31(1), 96-102. doi:10.1111/j.1468-0270.2010.02057.x.
- Berman, M., Miura, M., & Bergstresser, J. (2013). *Tobacco product display restrictions*. Retrieved from Boston, MA.
- Cancer Research, UK. A response by cancer research UK to Canada's ruinous tobacco display ban: Economic and public health lessons by Patrick Basham, Institute for Economic Affairs.
- Carter, O.B.J., Mills, B.W. & Donovan, R.J. (2009). The effect of retail cigarettes pack displays on unplanned purchases: results from immediate post-purchase interviews. *Tobacco Control*, 18, 218-221.
- Clattenburg, E. J., Elf, J. L., & Apelberg, B. J. (2013). Unplanned cigarette purchases and tobacco point of sale advertising: A potential barrier to smoking cessation. *Tobacco Control*, 22(6), 376. doi:10.1136/tobaccocontrol-2012-050427
- Cohen, J. E., Planinac, L. C., Griffin, K., Robinson, D. J., O'Connor, S. C., Lavack, A., Di Nardo, J. (2008). Tobacco promotions at point-of-sale. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique*, 99(3), 166-171.
- Cohen, J. E., Planinac, L., Lavack, A., Robinson, D., O'Connor, S., & DiNardo, J. (2011). Changes in retail tobacco promotions in a cohort of stores before, during, and after a tobacco product display ban. *American Journal of Public Health*, 101(10), 1879-1881. doi:10.2105/AJPH.2011.300172.
- Curry, L. E., Schmitt, C. L., & Juster, H. (2014). The Haverstraw experience: The first tobacco product display ban in the United States. *American Journal of Public Health*, 104(6), e9-e12. doi:10.2105/AJPH.2013.301861.
- Feighery, E.C., Henriksen, L., Wang, Y., Schleicher, N.C. & Fortmann, S. P. (2006). An evaluation of four measures of adolescents' exposure to cigarette marketing in stores. *Nicotine & Tobacco Research*, 8(16), 751-759.
- Germain, D., McCarthy, M., & Wakefield, M. (2010). Smoker sensitivity to retail tobacco displays and quitting: A cohort study. *Addiction*, 105(1), 159-163. doi:10.1111/j.1360-0443.2009.02714.x.
- Hastings, G., Mackintosh, A. M., Holme, I., Davies, K., Angus, K., & Moodie, C. (2008). *Point of sale display of tobacco products*. Retrieved from Stirling, UK.

- Hoek, J., Gifford, H., Pirikahu, G., Thomson, G., & Edwards, R. (2010). How do tobacco retail displays affect cessation attempts? Findings from a qualitative study. *Tobacco control, 19*(4), 334-337.
- Jaine, R., Russell, M., Edwards, R., & Thomson, G. (2014). New Zealand tobacco retailers' attitudes to selling tobacco, point-of-display bans and other tobacco control measures: A qualitative analysis. *New Zealand Medical Journal, 127*(1396).
- Johns, M., Sacks, R., Rane, M., & Kansagra, S. (2013). Exposure to tobacco retail outlets and smoking initiation among New York City adolescents. *Journal of Urban Health, 90*(6), 1091-1101. doi:10.1007/s11524-013-9810-2.
- Li, L., Borland, R., Yong, H-H., Sirirassamee, B., Hamman, S., Omar, M. & Quah, A.C.K. (2015). Impact of Point-of_Sale Tobacco Display Bans in Thailand: Findings from the International Tobacco Control (ITC) Southeast Asia Survey. *International Journal of Environmental Research and Public Health, 12*,9508-9522.
- MacGregor, J. (2008). *Tobacco advertising at point-of-sale*.
- MacKintosh, A. M., Moodie, C., & Hastings, G. (2012). The association between point-of-sale displays and youth smoking susceptibility. *Nicotine & Tobacco Research, 14*(5), 616. doi:10.1093/ntr/ntr185.
- Monshouwer, K., Verdumen, J., Ketelaars, T., & van Laar, M. (2104). *Points of sale of tobacco products: Synthesis of scientific and practice based knowledge on the impact of reducing the number of points of sale and restrictions on tobacco product displays*. Retrieved from Utrecht.
- Paynter, J., Edwards, R. & Schluter, P.J. (2009). Point of sale tobacco displays and smoking among 14-15 year olds in New Zealand: a cross-sectional study. *Tobacco Control, 18*, 268-274.
- Paynter, J. & Edwards, R. (2009). The impact of tobacco promotion at the point of sales: a systematic review. *Nicotine & Tobacco Research, 11*(1), 25-35.
- Pereira, I., & Veludo-de-Oliveira, T. M. (2014). Analysis of the promotion of cigarettes at the point of sale and its attractiveness to children. *International Journal of Consumer Studies, 38*(3), 297-306. doi:10.1111/ijcs.12099.
- Robertson, L., McGee, R., Marsh, L., & Hoek, J. (2014). A systematic review on the impact of point-of-sale tobacco promotion on smoking. *Nicotine & Tobacco Research*. doi:10.1093/ntr/ntu168.
- Schmitt, C. L., Allen, J. A., Kosa, K. M., & Curry, L. E. (2015). Support for a ban on tobacco powerwalls and other point-of-sale displays: Findings from focus groups. *Health Education Research, 30*(1), 98. doi:10.1093/her/cyu046.
- Thomson, G., Hoek, J., Edwards, R. & Gifford, H. (2008). Evidence and arguments on tobacco retail displays: marketing an addictive drug to children. *The New Zealand Medical Journal, 121*(1276), 87-98.

- Wakefield, M., Germain, D., Durkin, S. & Henriksen, L. (2006). An experimental study of effects on schoolchildren of exposure to point-of-sale cigarette advertising and pack displays. *Health and Education Research*, 21(3), 338-347.
- Wakefield, M., Germain, D. & Henriksen, L. (2008). The effect of retail cigarette pack displays on impulse purchase. *Addiction*, 103, 322-328.
- Wakefield, M., Zacher, M., Scollo, M., & Durkin, S. (2012). Brand placement on price boards after tobacco display bans: A point-of-sale audit in Melbourne, Australia. *Tobacco Control*, 21(6), 589-592. doi:10.1136/tobaccocontrol-2012-050616

ANNEX THREE – QUALIFICATIONS, RESUME AND SELECTED PUBLICATIONS

Biography

David Midgley joined INSEAD in 1999 as Professor of Marketing. Previously he held positions at the Anderson School, University of California, Los Angeles and the Australian Graduate School of Management, Sydney, where he was Foundation Chair. From 2001 to 2005 he was elected Area Head for the Marketing Area at INSEAD.

Professor Midgley is a graduate in science, management and marketing from the Universities of Salford and Bradford in England. He has been an invited scholar at Stanford Graduate School of Business and has also taught at the Wharton School, University of Pennsylvania.

Professor Midgley has over 120 publications, including papers in leading journals such as the *Journal of Consumer Research*, *Journal of Information Technology*, *Journal of International Business*, *Journal of Marketing Research*, *Journal of Marketing*, *Marketing Science*, *Management Science*, *Organization Science* and *Research Policy*. He has also written a number of books, including the *The Innovation Manual* and most recently, *Strategic Marketing for the C-Suite*. He has served on the editorial boards of the *Academy of Management Perspectives*, the *Journal of Consumer Research* and the *International Journal of Research in Marketing*. His principal areas of research are innovation and strategy. Other scholars have cited his work over 5,000 times.

Professor Midgley was Research Director for the Australian Federal Government's *Industry Task Force on Leadership and Management Skills* (the Karpin Committee). He is also author of major reports on management development for the Australian National Training Board and international competitiveness for Austrade. He has extensive consulting experience in Australia, Europe and North America.

At INSEAD Professor Midgley directs executive programmes for companies such as *Pernod Ricard* (France) and *Syngenta* (Switzerland). Over 2,000 managers and executives have participated in these programs. In 2015 his work with Syngenta received a best practice award from the European Foundation for Management Development. In the MBA programme, he teaches the *Market Driving Strategies* elective, which over 1000 students have attended since David joined INSEAD.

Citations

Total of 5,123 citations at February 21st, 2016 (Google Scholar). h-index 26, i10-index 43. 2,137 citations since 2011.

Top ten scientific papers

Citations	Title	Year	Journal
1,222	Innovativeness: the concept...	1978	Journal of Consumer Research
429	Formative versus reflective...	2008	Journal of Business Research
310	The nature of lead users...	2004	Research Policy
245	Patterns of interpersonal ...	1983	Journal of Marketing Research
238	A longitudinal study of product ...	1993	Journal of Consumer Research
208	Modeling asymmetric competition...	1988	Marketing Science
178	Dual paths to performance...	2005	Journal of International Business
140	Breeding competitive strategies...	1997	Management Science
134	The effects of network structure...	1992	Research Policy
123	The optimal performance of the...	2000	Organization Science

Citations to most recent scientific paper

1	Mindscales across landscapes...	2015	Journal of International Business
---	---------------------------------	------	-----------------------------------

Most cited practitioner article

260	Knowledge management...	2002	California Management Review
-----	-------------------------	------	------------------------------

Academic Positions

Standing Faculty

1. Professor of Marketing, INSEAD, 1999-
2. Foundation Chair of Marketing, Australian Graduate School of Management, 1987-1998.
3. Associate Professor, Australian Graduate School of Management, University of New South Wales, 1983-1987.
4. Senior Lecturer, Australian Graduate School of Management, University of New South Wales, 1979-1983.
5. Lecturer, Australian Graduate School of Management, University of New South Wales, 1976-1979.
6. Lecturer, School of Marketing, University of New South Wales, 1975-1976.
7. Research and Senior Research Fellow, Cranfield School of Management, 1972-1975.
8. SSRC Research Fellow, University of Bradford, 1971-1972.

Visiting Faculty

1. Alliance Exchange Professor, Wharton School, University of Philadelphia, 2003.
2. Visiting Professor, Anderson Graduate School of Management, UCLA, 1999.
3. Visiting Professor, INSEAD, 1998.
4. Visiting Professor, Anderson Graduate School of Management, UCLA, 1988, 1990, 1991, 1992, 1994 and 1995.
5. Visiting Associate Professor, Graduate School of Management, UCLA, 1980-1981, 1983 and 1987.
6. Visiting Scholar, Graduate School of Business, Stanford University, 1978.

Administrative positions

1. Area Chair, INSEAD Marketing Area, 2001-2005.
2. Director, Centre for Corporate Change, Australian Graduate School of Management, 1998-1999. Australian Research Council Special Centre.
3. Research Director, Industry Task Force on Management and Leadership Skills, Australian Federal Government, 1992-1995.
4. Head of the Marketing Group, Australian Graduate School of Marketing, 1987-1996.
5. Director, Centre for Export Marketing, Australian Graduate School of Management, 1987-1995.

Professional Service

1. Editorial Board, *Academy of Management Perspectives*, 2012-2015. Served as judge for 2012, 2013 & 2014 Best Paper Awards.
2. Editorial Board of the *International Journal of Research in Marketing* (North-Holland), 1990 to 1992, 2001-2006.
3. Editorial Board of the *Journal of Consumer Research* (University of Chicago Press) 1983-1999.
4. Occasional reviewer for the *Academy of Management Perspectives*, *Australian Journal of Management*, *California Management Review*, *International Journal of Forecasting*, *Internet Marketing and Advertising*, *Journal of Artificial Societies and Social Simulations*, *Journal of Management Studies*, *Journal of Marketing*, *Journal of Marketing Research*, *Journal of Organizational Behavior*, *Journal of Product Innovation Management*, *Journal of World Business*, *Long Range Planning*, *Management Science*, *Marketing Science*, *Research Policy*, *Schmalenbach Business Review* and *Technovation*. Reviewer for the AIB Annual Meeting, 2013. Reviewer for the *Annual Conference of the European Marketing Academy*, 1995 and 1999. Reviewer for the *Handbook of Consumer Theory and Research*, 1990. Member of the Editorial Board of the *Markstrat Newsletter*, 1988-1991. Member of the Editorial Board for the *Handbook of Marketing*, 2002.
5. Co-chair, *Enterprise Management in the Asia-Pacific Region*, New Delhi, 1996. Review Chair, *TIMS Marketing Science Conference*, Sydney, 1995. Workshop Member, *Banff Invitational Symposium on Consumer Decision-making and Choice*, 1990. Deputy Chairman, *TIMS XXVII International Conference*, Gold Coast, 1986.

Institutional Service

At INSEAD

1. Chair, Review Committee for the INSEAD Blue Ocean Strategy Institute, 2016.
2. EDP Committee, 2013-.
3. Chair, affiliate review committee, Tom Mannarelli, 2010.
4. Member, Research Productivity Taskforce, 2003.
5. Rapporteur, tenure case, Ziv Carmon, 2002.
6. Annual review of non-tenured faculty, 2002.
7. Area Chair, Marketing Area, 2001-2005.
8. Ad Hoc Committee, appointment of Miklos Sarvary, 2001.
9. Annual review of non-tenured faculty, 2001.
10. Rapporteur, promotion to Associate Professor, Prashant Malaviya, 2000.
11. Chair, Recruitment, Marketing Area, 1999-2001.

12. Member, MBA Committee, 1999-2001.

At AGSM, UNSW

13. Chair, Strategic Review, 1998.

14. Member, Review Committee, Institute for Research into International Competitiveness, Curtin University of Technology, 1997.

15. Chair, Committee to review the reappointment of the Dean, 1994.

16. Elected Presiding Member of Faculty, 1992-1994 and 1994-1996.

17. Chairman, Research Centre Policy Committee, 1990.

18. Director, MBA Program, 1989.

19. Member, MBA Redesign Committee, 1989.

20. Author, Review of MBA Administration, 1989.

21. Co-author, Review of Marketing, 1989.

22. Chairman, Computer Replacement Committee, 1989.

23. Co-author, AGSM Research Management Plan, 1988.

24. Chairman, Computer Committee, 1984-1986.

25. Elected Member, Board of Management, 1983-1984.

26. Co-author, Survey of UNSW Alumni, 1983.

27. Director, MBA Program, 1981-1983.

28. Member, Advisory Committees, Mitchell College, Bathurst, 1981-1986.

29. Member, Computer Committee, 1981-1986.

30. Member, Higher Degrees Committee, 1981-1987.

31. Member, Executive Program Advisory Committee, 1977-1979.

At Cranfield Institute of Technology

32. Member, Research Committee, 1973-1974.

At University of Bradford

33. Member, New Library Planning Committee, 1970-1971.

Doctoral Theses

Committee Chair

1. Venaik, S., (1999). *A New Model of Global Marketing*. Associate Professor, University of Queensland.

2. Morrison, P.D., *An Integrated Framework for Understanding the Industrial Adoption Decision* (1996). Full Professor, University of New South Wales.
3. Dawes, P.L., *Organizational Information Search When Purchasing Advanced Technology* (1989). Full Professor, University of Wolverhampton.
4. Wicking, B., *Credibility, Attraction and Power Effects in Interpersonal Exchange* (1982). Managing director, private sector.
5. Dowling, G.R., *Consumer Risk Handling Strategies* (1981). Full Professor, University of New South Wales.

Committee Member

1. Watson, Noshua, *Decoupling Compliance and Convergence: Strategic Responses to Institutional Pressures and Organizational Learning from Performance Feedback in the Case of Corporate Social Responsibility* (2009). Research Fellow, Accenture.
2. Sellier, Anne-Laure, *The Mediating Role of Counter-argumentation in the Processing of Mismatching Affective and Cognitive Persuasive Appeals: Implications for Attitude Change and Positioning of Late Entrants* (2003). Associate Professor, HEC.
3. Sa Vinhas, A. *Dual Distribution Channels in Business-to-Business Marketing: A Transaction Interdependencies View* (2002). Assistant Professor, Washington State University.
4. Soo, C. W., *The Process of Knowledge Creation in Organizations: An Empirical Investigation* (1999). Assistant Professor, University of Western Australia.
5. Hackman, B.K., *Managerial Delegation Behavior* (1990). HR Director, private sector.

Teaching

INSEAD Customized Executive Programmes

Directed programs for *Air Liquide* (France), *LG Electronics* (Korea), *Manpower* (USA), *Pernod Ricard* (France), *SSL International* (United Kingdom), *Syngenta* (Switzerland), *Umicore* (Belgium) and the *UAE Federal Government* (United Arab Emirates). Teach innovation and marketing strategy on other programs.

INSEAD MBA

The elective *Market Driving Strategies*.

Honours

The 1993 *Journal of Consumer Research* paper—a longitudinal study of product form innovation: the interaction between predispositions and social messages—(with Dowling) won the *1997 Australian and New Zealand Marketing Educators Distinguished Marketing Researcher Award*. This award is for the best paper published in an international journal by a scholar working in Australia in the five preceding years (1992 through 1996).

Publications

Working Papers

1. Marketing Strategy in MNC Subsidiaries: Pure versus Hybrid Archetypes (with Venaik). INSEAD Working Paper, 2012/12/131/MKT (Revision 2012/72/MKT), December 2012. Featured in the SSRN Top 10 downloads.
2. The Retailer Needs to be Paid for Sophisticated Decisions: Modeling Promotional Interactions Between Consumers, Retailers and Brand Managers (with Marks). On the advice of colleagues, being revised to incorporate additional objective functions into the model and extend the Monte Carlo simulations. INSEAD Working Paper 2008/65/MKT.

Refereed Journal Articles

1. Mindscapes across landscapes: transnational and subnational culture archetypes (with Venaik). *Journal of International Business Studies*, 46(9), 2015, 1051-1079.
2. Strategic marketing for the C-suite: a review of the research literature and its relevance to senior executives. *Foundations and Trends in Marketing*, 8(3-4), 2013, 147-341.
3. CRM and Firm Performance (with Coltman and Devinney). *Journal of Information Technology*, 26, 2011, 205-219.
4. The value of managerial beliefs in turbulent environments: managerial orientation and e-business advantage (with Coltman and Devinney), *Journal of Strategy and Management*, 1 (2), 2008, 181-197.
5. Formative versus reflective measurement models: two applications of formative measurement (with Coltman, Devinney and Venaik), *Journal of Business Research*, 61 (12), 2008, 1250-1262.
6. The building and assurance of agent-based models: an example and challenge to the field (with Kunchamwar and Marks), *Journal of Business Research*, 60 (8), 2007, 884-893.
7. External knowledge acquisition, creativity and learning in organizational problem solving (with Devinney and Soo), *International Journal of Technology Management*, 38 (1/2), 2007, 137-159.
8. E-business strategy and firm performance: a latent class assessment of the drivers and impediments to success (with Coltman and Devinney), *Journal of Information Technology*, 22 (2), 2007, 87-101. Lead article.
9. Organizational learning in high-technology purchase situations: the antecedents and consequences of the participation of external IT consultants (with Dawes and Lee), *Industrial Marketing Management*, 36 (3), 2007, 285-299.
10. Dual paths to performance: the impact of global pressures on MNC subsidiary conduct and performance (with Devinney and Venaik), *Journal of International Business Studies*, 36 (6), 2005, 655-675.

11. A New Perspective on the Integration-Responsiveness Pressures Confronting Multinational Firms (with Devinney and Venaik), *Management International Review*, 44(1), 2004, 17-50.
12. The Nature of Lead Users and Measurement of Leading Edge Status (with Morrison and Roberts), *Research Policy*, 33 (2), 2004, 351-362.
13. What to Codify: Marketing Science or Marketing Engineering, *Marketing Theory*, 2(4), 2002, 363-368.
14. Knowledge management: philosophy, process, pitfalls, and performance, *California Management Review*, (with Soo, Devinney and Deering), 44(4), 2002, 129-150.
15. Keeping E-business in perspective (with Coltman, Devinney and Latukefu), *Communications of the Association for Computing Machinery*, 45(8), 2002, 69-73.
16. An empirical assessment of e-business implementation Constraints (with Coltman and Devinney), *The e-Business Review*, 2, 2002, 68-71.
17. E-business: revolution, evolution or hype? (with Coltman, Devinney and Latukefu), *California Management Review*, 44(1), 2001, 57-86.
18. The organizational imperative and the optimal performance of the global firm: formalizing and extending the integration-responsiveness framework (with Devinney and Venaik), *Organization Science*, 11(6), 2000, 674-695.
19. Opinion leadership amongst leading edge users (with Morrison and Roberts), *Australasian Marketing Journal*, 8(1), 2000, 5-14. *This paper won the Australian and New Zealand Marketing Academy's Roger Layton Best Paper Award.*
20. The complexity of competitive marketing strategies (with Marks, Cooper and Shiraz), *Complexity International*, 6, 1999, <http://life.csu.edu/complex/ci/vol6>.
21. Breeding competitive strategies (with Marks and Cooper), *Management Science*, 43(3), 1997, 257-275. *Lead article.*
22. Factors Affecting the Involvement of Technical Consultants in High Technology Business Markets (with Dawes and Patterson), *Journal of Business and Industrial Marketing*, 12(2), 1997, 83-102. *Lead article.*
23. The decision processes of innovative communicators and other adopters (with Dowling), *Marketing Letters*, 4(4), 1993, 297-308.
24. A longitudinal study of product form innovation: the interaction between predispositions and social messages (with Dowling), *Journal of Consumer Research*, 19(4), 1993, 611-625. Abstracts of this paper were also published in the *Journal of Product Innovation Management*, 10(5), 1993, 439 and the *Digest of Management Research*, 4(4), 1993, 51-53.
25. Comment on going international by Yetton, Davis and Swan, *Australian Journal of Management*, 17(1), 1992, 169-173.

26. The effect of network structure in industrial diffusion processes (with Morrison and Roberts), *Research Policy*, 21(6), 1992, 533-552.
27. Using rank values as an interval scale (with Dowling), *Psychology and Marketing*, 8(1), 1991, 37-42.
28. Asymmetric market share models (with Carpenter, Cooper and Hanssens), *Marketing Science*, 7(4), 1988, 393-412.
29. The coarse and fine structure of data (with Dowling), *Decision Sciences*, 19(4), 1988, 830-847.
30. Product evaluations in a dynamic market (with Dowling), *Psychology and Marketing*, 3(2), 1986, 99-111.
31. The case for user-oriented innovation: comments on the National Technology Strategy (with Roberts), *Australian Journal of Management*, 9(2), 1984, 43-58.
32. Parsimony or explanation: on the estimation of systems defined by nonlinear differential equations, *Journal of Consumer Research*, 10(4), 1984, 445-448.
33. Patterns of interpersonal information seeking for the purchase of a symbolic product, *Journal of Marketing Research*, 20(1), 1983, 74-83. An abstract of this paper was also published by the *Journal of Marketing*, 47(4), 1983, 137.
34. Toward a theory of the product life cycle: explaining diversity, *Journal of Marketing*, 45(4), 1981, 109-115.
35. The decision process models of Antarctic travel innovators (with Dowling), *Australian Journal of Management*, 3(2), 1978, 147-162.
36. Innovativeness: the concept and its measurement (with Dowling), *Journal of Consumer Research*, 5(4), 1978, 229-242.
37. A simple mathematical theory of innovative behavior, *Journal of Consumer Research*, 3(1), 1976, 31-41.
38. The overlap problem, *European Journal of Marketing*, Summer 1970, 165-175.
39. Institutional influence in air passenger routing decisions (with Wills), *European Journal of Marketing*, Winter 1969, 259-266.

Conference Proceedings

40. Marketing strategy configurations: pure versus hybrid archetypes (with Venaik), *Proceedings of the AIB Annual Meeting*, Istanbul, July 2013.
41. Marketing management in MNC subsidiaries: an archetypal analysis (with Venaik), *Proceedings of the AIB Annual Meeting*, Washington, July 2012.
42. Embodied agents on a branding website: Deepening on a website stickiness through an attitudinal persuasion route (with Chandon and Diesbach), *8th Nantes Conference on E-marketing*, Nantes: September 2008.

43. Embodied agents on a branding website: deepening website stickiness through an attitudinal persuasion route (with Diesbach), *Proceedings of the 37th EMAC Conference*, Brighton, May 2008.
44. Embodied agents on a communication and branding website: modeling effects through an affective persuasion route (with Diesbach), in C. Acevedo, J. Hernandez and T. Lowrey (Eds.), *Latin American Advances in Consumer Research Volume 2*, Sao Paulo, Brazil: Association for Consumer Research, August 2008.
45. Embodied agents on commercial websites: modeling their effects through an affective persuasion route (with Diesbach), in H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerstahl and P. Ohrstrom (Eds.), *Persuasive Technology 2008*, Lecture Notes in Computer Science 5033, Berlin: Springer-Verlag, 2008, 283-286.
46. Embodied agents on a website: modeling an attitudinal route of influence (with Diesbach), in Y. Kort, W. Ijsselsteijn, C. Midden, B. Eggen and B. Fogg (Eds.), *Persuasive Technology 2007*, Lecture Notes in Computer Science 4744, Berlin: Springer Verlag 2007, 223-230.
47. Embodied virtual agents: an affective and attitudinal approach of the effects on man-machine stickiness in a product/service discovery (with Diesbach), in D. Harris (Ed.), *HCI International 2007*, Lecture Notes in Artificial Intelligence 4562, Berlin: Springer Verlag, 2007, 42-51.
48. Firms as seminaries: Environmental and organizational influences on learning, innovation and performance (with Devinney and Venaik), in L.W. Malmskov & K.S. Jacobsen (editors), *The MNC as a Knowing Organization: 29th Annual EIBA Conference Proceedings*, Copenhagen Business School, Copenhagen, Denmark, December 2003.
49. Searching for markov perfect equilibria in oligopolistic price wars (with Marks and Cooper), *The Industry Economics Conference*, Sydney, July 10-11, 2000.
50. International perspectives on the state of the E-business revolution (with Coltman, Devinney and Latukefu), *ECommerce and Global Business Forum: Building Electronic Bridges across Nations*, edited by J. de la Torre and R. Moxon, Santa Cruz, California, May 2000.
51. Co-evolution with the genetic algorithm: application to repeated differentiated oligopolies (with Marks, Cooper and Shiraz), *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, edited by W. Banzhaf, J. Daida, A.E. Eiben, M.H. Garzon, V. Honavar, M. Jakiela and R.E. Smith , Orlando, Florida, July 13-17, 1999, 1609-1615.
52. Using genetic algorithms to breed competitive marketing strategies (with Marks, Cooper and Shiraz), in the *Proceedings of the 1998 IEEE International Conference on Systems, Man and Cybernetics*, La Jolla, CA, October 11-14 1998. *Intelligent Systems for Humans in the Cyberworld*, New York: IEEE 2367-2372.
53. The complexity of competitive marketing strategies (with Marks, Cooper and Shiraz), *Complex Systems '98: Complexity between the Ecos—from ecology to*

economics, Proceedings of the Fourth Conference on Complex Systems, edited by R. Standish, B. Henry, S. Watt, R. Marks, R. Stocker, D. Green, S. Keen and T. Bossomaier, Sydney, UNSW November-December 1998, 336-345.

54. The role of leading edge users in the adoption of technological innovations by organizations (with Morrison and Roberts), *Asia Pacific Advances in Consumer Research*, Volume 1, Provo, Utah: ACR 1994.
55. Coffee price wars and the genetic algorithm: estimating strategies in a real-world oligopoly (with Marks and Cooper), *Emergent Structures*, Santa Fe: Santa Fe Institute, July 1993, 7-8.
56. The nature of communication networks between organizations in the diffusion of technological innovations (with Morrison and Roberts), in Johnson M. (ed.), *Advances in Consumer Research*, Volume 18, Provo, Utah: ACR 1990.
57. Cognitive and social influences in purchase behavior (with Morrison and Dowling), in Srull T. (ed), *Advances in Consumer Research*, Volume 16, Provo, Utah: ACR 1988.
58. A meta-analysis of the diffusion literature, in Anderson P.F. and M. Wallendorf (eds.), *Advances in Consumer Research*, Volume 14, Provo, Utah: ACR 1986.
59. Cross-cultural differences in information source importance (with Dowling), in Keown C.F. and A.G. Woodside (eds), *Proceedings of the 1984 Comparative Consumer Psychology Conference*, American Psychological Association, December 1984, 33-35.
60. Innovation in the male fashion market: The parallel diffusion hypothesis, *Proceedings of the 31st ESOMAR Seminar*, Barcelona: ESOMAR, December 1974.
61. Current progress in the application of a theory of innovative behavior, *Proceedings of the 27th ESOMAR Seminar*, Amsterdam: ESOMAR, January 1974.
62. Innovation in the fashion market, the 11th Marketing Theory Seminar, *Innovation in Marketing*, Strathclyde: University of Strathclyde, May 1973.

Books

63. *Strategic Marketing for the C-suite*, Hanover, MA: now Publishers, 2013. Monograph in the *Foundations and Trends in Marketing* series.
64. *The Innovation Manual: Integrated Strategies and Practical Tools for Delivering Value Innovation to the Market*, Chichester, UK: John Wiley & Sons, 2009.
65. *New Horizons: Enterprise Management in the Asia-Pacific Region* (edited with Neelamegham and Sen), New Delhi: Tata McGraw-Hill, 1998.
66. *Research Report of the Industry Task Force on Leadership and Management Skills -- Volume III of Enterprising Nation: Renewing Australia's Managers to Meet the Challenges of the Asia-Pacific Century* (edited with Mozell), Canberra: Australian Government Publishing Service, 1995.

67. *Australian Management and International Competitiveness*, Canberra: Australian Trade Commission, 1991.
68. *Benchmark Study of Management Development*, Commissioned Report Number 5, National Board of Employment, Education and Training, Canberra: Australian Government Publishing Service, 1990.
69. *Innovation and New Product Marketing*, New York: Wiley, London: Croom Helm, 1977.
70. *Management Problems in Fashion Industries* (with Wills), Cranfield: Cranfield Institute Press, 1975.
71. *Customers in Action* (with Christopher), Bradford: MCB Books, 1975.
72. *Fashion Marketing* (edited with Wills), London: George Allen and Unwin, 1973.
73. *Creating and Marketing New Products* (edited with Hayhurst and Wills), London: Crosby Lockwood Staples, 1973.

Chapters in Books

74. Better delivery of new innovation by doing three things at once, in Ton Langelier (editor) *Innopreneur*, New York: Channel V Books, 2012, 169-170.
75. Co-evolving better strategies in oligopolistic price wars (with Marks), in Jean-Philippe Renard (editor) *Handbook of Research on Nature Inspired Computing for Economy and Management*, Idea Group, 2006.
76. Strategy content and process in the context of e-business performance (with Coltman and Devinney), in Yves Doz, Gabriel Szulanski and Joe Porac (eds.), *Strategy Process, Volume 22 of Advances in Strategic Management*, Amsterdam: Elsevier, 2005, 349-386.
77. Relating customer value to strategic competence: a discrete choice measurement approach (with Brazell and Devinney), in Ron Sanchez & Jorg Freiling (editors), *the Marketing Process in Organizational Competence, Volume 1 of Research in Competence-Based Management*, Amsterdam: Elsevier, 2005, 15-45.
78. Knowledge creation in organizations: a multiple study overview (with Devinney and Soo), in Joseph Davis, Eswaran Subrahmanian and Art Westerberg (eds), *Knowledge Management: Organizational and Technological Dimensions*, Berlin: Physica-Verlag, 2005, 79-96.
79. ChateauOnline in Tawfik Jelassi and Albrecht Enders (editors), *Strategies for e-Business: Creating Value through Electronic and Mobile Computing*, Harlow: Pearson Education, 2004.
80. Dual paths to multinational subsidiary performance: networking to learning and autonomy to innovation (with Devinney and Venaik), in Africa Arino, Pankaj Ghemawat and Joan Ricart (eds), *Creating Value through International Strategy*, Basingstoke & New York: Palgrave MacMillan 2004, 130-144.
81. E-business performance: a latent class examination (with Coltman and Devinney), in G. Fandel, U. Backes-Gellner, M. Schluter and J.E. Staufenbiel (eds), *Modern*

Concepts of the Theory of the Firm: Managing Enterprises in the New Economy, Berlin: Springer, 2003, 58-68.

82. The role of knowledge quality in firm performance (with Soo and Devinney), in H. Tsoukas and N. Mylonopoulos (eds.), *Organizations as Knowledge Systems: Knowledge, Learning and Dynamic Capabilities*, Palgrave: London, 2003, 252-275.
83. Managerial beliefs, market contestability and dominant strategic orientation in the eclectic paradigm (with Devinney and Venaik), in Cantwell, J. and R. Narula (Editors), *International Business and the Eclectic Paradigm: Developing the OLI Framework*, London: Routledge, 2003, 152-173.
84. How will enterprises be competitive in the 21st century: Australian business response, in Midgley, D.F., S. Neelamegham and C. Sen (Editors), *New Horizons: Enterprise Management in the Asia-Pacific Region*, New Delhi: Tata McGraw-Hill, 1998.
85. Adaptive behavior in an oligopoly (with Marks), in Biethahn, J. and V. Nissen (Editors), *Evolutionary algorithms in management applications*, Berlin: Springer-Verlag, 1995, 225-239.
86. Design of the research program, in Midgley, D.F. and S.L. Mozell (Editors), *Enterprising Nation: Renewing Australia's Managers to Meet the Challenges of the Asia-Pacific Century (Volume III)*, Canberra: Australian Government Publishing Service, 1995.
87. The need for leadership and management skills, in Midgley, D.F. and S.L. Mozell (Editors), *Enterprising Nation: Renewing Australia's Managers to Meet the Challenges of the Asia-Pacific Century (Volume III)*, Canberra: Australian Government Publishing Service, 1995.
88. Community attitudes to small business (with Mozell), in Midgley, D.F. and S.L. Mozell (Editors), *Enterprising Nation: Renewing Australia's Managers to Meet the Challenges of the Asia-Pacific Century (Volume III)*, Canberra: Australian Government Publishing Service, 1995.
89. How can Australia improve: overall conclusions, in Midgley, D.F. and S.L. Mozell (Editors), *Enterprising Nation: Renewing Australia's Managers to Meet the Challenges of the Asia-Pacific Century (Volume III)*, Canberra: Australian Government Publishing Service, 1995.
90. Comments on Susan Douglas, et al, in Lilien, G., Laurent, G. and B. Pras (Editors), *Research Traditions in Marketing*, Dordrecht: North Holland 1993.
91. A simple mathematical theory of innovative behavior, in Wind, Y., Mahajan, V. and R. Cardozo (Editors), *New Product Forecasting*, Lexington: Lexington Books, 1981.
92. The management of fashion, *Cranfield Research Papers in Marketing and Logistics*, Number 9, Cranfield: Cranfield Institute Press, 1975.
93. Managing new products, *Cranfield Research Papers in Marketing and Logistics*, Number 8, Cranfield: Cranfield Institute Press, 1975.

Instructional Materials

94. *Bet on Micro Four Thirds A* (with Bart), INSEAD 2009 (Pre-Release Version)
95. *Chateau Online* (with Devinney and Snall), INSEAD 2003.
96. *Fly Buys™: the more you buy, the more you fly*, Melbourne: Melbourne Business School Case Library, 1996.
97. *Marketing Strategy*, Sydney: AGSM 1996-1998.
98. *Moldflow A & B* (with Jones, Marsh, and Mozell), Sydney: AGSM 1993.
99. *Marketing Principles* (with Dowling, Roberts and Rossiter), Sydney: AGSM, 1990-1998.
100. *Export Planning* (with Dommett), Canberra: AUSTRADE, 1989.
101. The product life cycle and new product marketing, in Le Claire, K. (Editor), *Marketing Management*, Geelong: Deakin University Press, 1983.
102. Product and service strategy, in Le Claire, K. (Editor), *Marketing Management*, Geelong: Deakin University Press, 1983.

Professional and Other Articles

103. Understanding customers in the solution economy, *Harvard Business Review Blog Network*, August 2012.
104. Gaming innovation (with De Pommes), *Chief Learning Officer*, January 2010, 30-31.
105. Knowledge management: eight key lessons for knowledge managers (with Soo, Devinney and Deering), *efmd Forum*, April 2002, 34-45.
106. The seven deadly fallacies of knowledge management (with Soo, Devinney and Deering), *Boss Magazine*, January 2001.
107. Your future: the expert files, *Professional Marketing*, March 1996, 10-11.
108. Schools need benchmarks based on their outcomes, *HR Monthly*, October 1995, 22-24.
109. International marketing, series of eight short articles (with Dowling, Bleasel and Sivashankar), *Trade Intelligence Report*, Canberra: Austrade, February through September 1993.
110. Market-driven innovation: a solid foundation for export (with Dowling), *AIM Magazine*, August 1992, 17-18.
111. The international competitiveness of Australian enterprises and the role of government policy in industry development, *Commonwealth Bank Lecture*, Australian Chamber of Commerce, 1990.
112. Forecasting the sales lifecycle for new biotechnological products, *Australian Biotechnology*, 2, 2 (October), 1988, 137-140.

113. Customer-oriented strategies for new biotechnological products, *Australian Biotechnology*, 1, 4 (March) 1988, 15-18.
114. The prospects for L-Lysine production in Australia (with Rogers, Cail and Fryer), *Journal of Food Technology*, November 1986.
115. Management in the nineties, the *Catalyst* (Australian Institute of Systems Analysts), January 1986, 1-4.
116. The evolution of market research in the next decade, *Research News* (Australian Market Research Society), 1(4), December 1984, 15-16.
117. Academia examines the role of management schools in forwarding marketing craft (with Roberts), *Marketing World*, (Australian Marketing Institute), 2(1), October 1984.
118. What makes people buy fashion, *Apparel Production and Marketing*, 3, June 1972, 7-8.
119. Register of research in the UK and Eire, *European Journal of Marketing*, Summer 1971.
120. Conspectus of advanced marketing courses, *European Journal of Marketing*, Winter 1970.

Reports

121. The strategic application of e-intelligence (with Coltman and Devinney), *Centre for Corporate Change, Australian Graduate School of Management*, 2001.
122. Report on the survey of Australian management and international competitiveness, *Office of Multicultural Affairs*, Department of Prime Minister and Cabinet, Commonwealth of Australia, 1991.
123. The role of technical and further education in management development, *National Board of Employment, Education and Training*, 1991.
124. Forecasting and adoption of new high technology products (with Morrison and Roberts), *Australian Telecommunications and Electronics Research Board*, 1989.

Conference & Other Academic Papers Presented

1. The influence of organizational structure on international purchasing success (with Tressin and Richter). *2nd International Symposium on Partial Least Squares*, Seville: June 2015.
2. The effects of virtual agents on customer usage of commercial websites (with Diesbach, Kuisma and Simola), *Seminar on Consumer Behavior in the Information Society*, Aalto University School of Economics, Helsinki, Finland, May 2011.
3. Managing the transition from product to service innovation (with Coltman), *The Auckland Workshop on Service-Dominant Logic*, Auckland, New Zealand, March 2011.

4. Towards multidisciplinary models of diffusion: phenomena to include and approaches to validation, *Workshop on Agent-Based Simulation of Diffusion Processes*, University of Vienna, Vienna, Austria, April 2010.
5. Issues in validating agent-based models using extant data (with Klapper and Marks), *2009 NAACSOS Annual Conference*, Arizona State University, Tempe, Arizona, USA, October 2009.
6. Embodied agents on a communication and branding website: modeling effects through an affective persuasion route (with Diesbach), *Latin American Advances in Consumer Research*, Sao Paulo, Brazil, July 2008.
7. Embodied agents on commercial websites: modeling their effects through an affective persuasion route (with Diesbach), *Persuasive Technology 2008*, Oulu, Finland, June 2008.
8. Embodied virtual agents: an affective and attitudinal approach of the effects on man-machine stickiness in product/service discovery (with de Diesbach), *Human Computer Interaction International 2007* conference, Beijing, China, July 2007.
9. Embodied agents on a website: modeling an attitudinal route of influence (with Diesbach), in Y. Kort, W. Ijsselsteijn, C. Midden, B. Eggen and B. Fogg (Eds.), *Persuasive Technology 2007*, Palo Alto, CA, April 2007.
10. Minimalism and model-building: an assured model of the exchanges between consumers, retailers and manufacturers (with Klapper, Kunchamwar and Marks), *5th Lake Arrowhead Conference on Human Complex Systems*, Lake Arrowhead, California, USA, April 2007.
11. Assuring Agent-Based Models: the 'Next Practice' in Modeling Diffusion (with Marks), *Accelerating Market Acceptance in a Networked World*, Marketing Science Institute/USC Conference, Los Angeles, CA, March 2007.
12. Marketing mix innovation in MNC subsidiaries: the impact of environmental and organizational factors (with Devinney and Venaik), *New Perspectives on Multinational Subsidiary Research Conference*, University of Newcastle, Newcastle, Australia, November 2006.
13. Destructive testing and empirical validation of a multi-agent simulation of consumers, retailers and manufacturers (With Marks and Kunchamwar), *3rd Lake Arrowhead Conference on Human Complex Systems*, Lake Arrowhead, California, USA, May 2005.
14. Development and validation of a multi-agent simulation of consumers, retailers and manufacturers (With Marks and Kunchamwar), *Agent-based Models of Market Dynamics and Consumer Behavior*, University of Surrey, Guildford, United Kingdom, January 2006.
15. The process of knowledge creation in organizations (With Soo and Devinney), *Academy of Management Annual Conference*, Hawaii, USA, August 2005.
16. Integrating and relating constructs used to study organizational adoption of technological innovations (With Morrison and Roberts), *Anzmac/EMAC Symposium*, Milan, Italy, May 2005.

17. The interaction between consumers, retailers and manufacturers: development of an empirically grounded and testable multi-agent simulation (with Marks and Kunchamwar), *2004 Lake Arrowhead Conference on Human Complex Systems*, Lake Arrowhead, California, USA, May 2004.
18. Development of an empirically grounded and testable multi-agent simulation of consumers, retailers and manufacturers (with Marks and Kunchamwar), *EUMAS 2004*, Barcelona, Spain, December 2004.
19. The interaction between consumers, retailers and manufacturers: development of an empirically grounded and testable multi-agent simulation (with Marks and Kunchamwar), *INFORMS Marketing Science Conference 2004* Rotterdam, The Netherlands, June 2004.
20. Knowledge from external sources: the impact on creativity and learning (with Soo and Devinney), *ANZAM*, Dunedin, New Zealand, December 2004.
21. The sellers and the store: explorations of a moderated oligopoly (with Marks), *2003 Lake Arrowhead Conference on Human Complex Systems*, Lake Arrowhead, California, March 2003.
22. Networking and inter-unit learning in a model of MNC subsidiary innovation and performance (with Venaik and Devinney), *International Workshop on Creating Value Through Global Strategy*, IESE Business School, Barcelona, Spain, June 2003.
23. Strategic drivers and organizational impediments to firm returns from IT investments: applications of conjoint analysis and best-worst scaling to identifying firm specific solutions (with Coltman and Devinney), *SAP Innovation Congress EMEA '03*, Basel, Switzerland, September 2003.
24. A new dynamic model integration-responsiveness in multinational firms (with Venaik and Devinney), *Annual Meeting of the Academy of International Business*, San Juan, June 2002.
25. Endogenizing the actions of artificial brand managers (with Marks and Cooper), *Agent-Based Modeling in the Social Sciences*, Lake Arrowhead, California, May 2002.
26. An empirical assessment of winners and losers in an e-enabled world (with Coltman and Devinney), *2002 Annual Conference of the International Academy of E-Business*, Orlando, Florida, March, 2002.
27. Knowledge management (with Soo and Devinney), *ANZAM Annual Conference*, Auckland, December 2001. Awarded "Highly Commended Paper."
28. An empirical assessment of e-business implementation: organizational and managerial constraints to gaining advantage through the internet (with Coltman and Devinney), *eLab, INSEAD & The Center for Management in the Information Economy, UCLA Research Conference*, Singapore, October 2001.
29. The relationship of lead users to the adopter categories used in marketing (Morrison and Roberts), *INFORMS Marketing Science Conference*, Wiesbaden, July 2001.

30. Modeling the dynamics of strategic choice: a preliminary investigation into the role of managerial beliefs and transactional structures on e-business orientation (with Coltman and Devinney), *Management in the Information Economy*, Anderson School, UCLA, February 2001.
31. An empirical examination of the dimensionality of the integration-responsiveness framework (with Venaik and Devinney), *In Search Of Relevance for International Business Research: Impact on Management and Public Policy*, AIB 2000 Annual Meeting, Phoenix, Arizona, November 2000.
32. Knowledge creating processes and innovation (with Soo and Devinney), *Strategy in the Entrepreneurial Millenium: Strategic Management Society's 20th Annual International Conference*, Vancouver, October 2000.
33. Refining the breeding of hybrid strategies: repeated games in the coffee market (with Cooper and Marks), *Computational Modeling Workshop*, Anderson School, UCLA, 1999.
34. An alternative methodology for the analysis of MNE performance (with Roth, Devinney and Venaik), *Annual Meeting of the Academy of International Business*, Vienna, October 1998.
35. Towards a finer understanding of lead users (with Roberts and Morrison), *INFORMS Marketing Science Conference*, Fontainebleau, July 1998.
36. The international marketing program in multinational firms: an empirical investigation using the extended integration-responsiveness framework (with Venaik and Devinney), *INFORMS Marketing Science Conference*, Fontainebleau, July 1998.
37. Modeling the three c's: a multiple stated choice experiment gap analysis approach (with Brazell and Devinney), *INFORMS Marketing Science Conference*, Fontainebleau, July 1998.
38. Breeding competitive strategies—recent extensions (with Cooper and Marks), *INFORMS Fall Conference*, Dallas, TX, October 1997.
39. Refining the breeding of hybrid strategies (with Marks and Cooper), *Third International Conference on Computing in Economics and Finance*, Stanford, CA, July 1997.
40. Validity in management training with reference to the Karpin report, *The Academy of the Social Sciences of Australia*, Melbourne, Victoria, July 1997.
41. Breeding competitive strategies extended (with Cooper and Marks), *INFORMS Marketing Science Conference*, Berkeley, California, March 1997.
42. How will enterprises be competitive in the 21st Century, *Enterprise Management in Asia Pacific Region*, New Delhi, October 1996.
43. Transactional and organisational imperatives in an integration - responsiveness framework (with Devinney and Venaik), *European Academy of International Business*, Stockholm, October 1996.

44. How boundedly rational are brand managers (with Marks and Cooper), *INFORMS Marketing Science Conference*, Sydney, July 1995.
45. Inter- and Intra-Firm Diffusion Patterns (with Roberts and Morrison), *TIMS XXXIII International Conference*, Singapore, June 1995.
46. Adaptive behavior in an oligopoly (with Marks and Cooper), *1994 Australasian Meeting of the Econometric Society*, Armidale, New England, July 1994.
47. International diffusion of a technological new product (with Devinney), *ORSA/TIMS Marketing Science Conference*, Tucson, Arizona, March 1994.
48. Modeling the use of consultants in high technology purchasing decisions (with Dawes and Patterson), *New Zealand Educators Conference*, Lincoln, New Zealand, December 1993.
49. Adaptive behavior in an oligopoly (with Cooper and Marks), *15th Annual conference of the Society for Economic Dynamics and Control*, Nafplion, Greece, June 1993.
50. Breeding optimal strategies for multiple period competitive games in mature markets (with Cooper and Marks), *ORSA/TIMS Marketing Science Conference*, St. Louis, Missouri, March 1993.
51. Coffee price wars and the genetic algorithm: estimating strategies in a real-world oligopoly (with Cooper and Marks), Invited Colloquium Series, *Santa Fe Institute*, February 1993.
52. Calibrating models of intra- firm and inter-firm diffusion (with Morrison and Roberts), *ORSA/TIMS Marketing Science Conference*, London, July 1992.
53. The role of leading edge users in the adoption process for industrial innovations (with Morrison), *ORSA/TIMS Marketing Science Conference*, London, July 1992.
54. Intra-and inter-organizational diffusion models (with Morrison and Roberts), *ORSA/TIMS Marketing Science Conference*, Wilmington, Delaware, March 1991.
55. The evolution of communication networks between leading edge organizations and their role in the diffusion of technological innovations (with Morrison and Roberts), *The Network of Innovators Conference*, Montreal May 1990.
56. The nature of communication networks between organizations involved in the diffusion of technological innovations (with Morrison and Roberts), *ORSA/TIMS Marketing Science Conference*, Urbana-Champaign, March 1990.
57. The role of applications in examining leading edge users in new technology forecasting (with Morrison and Roberts), *ORSA/TIMS Marketing Science Conference*, Duke University, March 1989.
58. Specific content versus general rules: expert systems for codifying market knowledge (with Dommett and McBurney), *ORSA/TIMS Marketing Science Conference*, Seattle, March 1988.

59. The role of expert systems in export marketing (with McBurney), *ORSA/TIMS Marketing Science Conference*, Paris, June 1987.
60. Simulating the properties of a biological model of brand competition (with J. Cooper, and Dowling), *ORSA/TIMS Marketing Science Conference*, Paris, June 1987.
61. A non-equilibrium theory of the selection of brand characteristics, *TIMS XXVII International Conference*, Gold Coast, July 1986.
62. Asymmetric market share models (with Carpenter, Cooper and Hanssens), *Joint Meeting of TIMS/ORSA*, Dallas, November 1984.
63. Segment differences in social, perceptual and preference change (with Dowling), *Management Educators Conference*, Melbourne, August 1982.
64. Centrality of beliefs and the impact of advertising (with Zielinski), *Management Educators Conference*, Melbourne, August 1982.
65. Relating similarity to preferences (with Cooper and Nakanishi), *Psychometric Society*, Montreal, May 1982.
66. Perceptions of fashion: issues of stability and change (with Cooper), Division 23, Annual Conference of the *American Psychological Association*, Los Angeles, August 1981.
67. Towards a theory of the product life cycle: prospects and problems, *Management Educators Conference*, Sydney, May 1978.

Industry Papers Presented

68. Generation Z: opportunities and challenges for marketing, *Teradata Universe*, Singapore, May 2010.
69. Frameworks and tools for value and business model innovation, *Front End of Innovation Europe*, Amsterdam, February 2010.
70. How competent are Australian managers, *Management in Australia—is it doing the job*, Securities Institute of Australia, August 1998.
71. Data warehousing and data mining: state of the art solutions, *NextDirect - Interactive*, Sydney, March 1997.
72. The 21st Century marketing organisation, *Marketing 96: Meeting the Challenge of Change*, Australian Marketing Institute National Conference, Sydney, September 1996.
73. Implications and market design of postgraduate courses, *Marketing and Strategic Planning in Education*, Sydney, April 1996.
74. How do Fly Buys, the GM Card and Telstra Visa compare competitively, *Card Wars II: The battle for loyalty continues*, Sydney, February 1996.
75. Line management's critical role in developing a learning organisation, *The Learning Organisation*, Sydney, July 1995.

76. Determinants of success in multi-partner loyalty schemes, *Card Wars: Customer Loyalty and Reward Programs*, Sydney, May 1995.
77. International Marketing, *AMP Insights Conference: Exploring Australia's Role in the Fastest Developing Region in the World*, Coolumb, February 1995.
78. Lessons from the Task Force Research, *Best Practice in Management Development Conference*, Melbourne, May 1994.
79. International marketing management: issues and ideas from Australian management, *International Business Week*, Sydney, October 1993.
80. Developing world competitive businesses," *First Annual Conference of the Processed Foods Industry*, Sydney, March 1991.
81. Australian communications markets and export infrastructures," *New Government Business Enterprises Conference*, IBC: Melbourne, August 1989.
82. Marketing renewable energy products internationally, *ERDIC/ANZSES Joint Seminar*, Sydney, July 1989.
83. Export strategies - how can small science-based enterprises break into major international markets, *ASIA Export Seminar*, Sydney, June 1989.
84. Management in the nineties, *Institute of Management Consultants*, Sydney, December 1986.
85. The Marketing of technology, *ERDIC Seminar*, Sydney, November 1985.
86. Getting it all together, *Joint Workshop of the NSW Innovation Centre and the Inventors Society*, Sydney, October 1984.
87. Forecasting consumer behavior in the fashion market (with Cantlay), *Yorkshire Textile Institute*, February 1973.

Industry Service

1. Academic Liaison Committee, *CMO Council*, 2012-.
2. Expert Reviewer, *UNEP Workshop on Eco-Innovation*, April 2013.
3. Member Advisory Committee, *BrandYou: a LinkedIn initiative*, 2010-2012.
4. Judge, *European Business Awards (Innovation Category)*, 2007.
5. Speaker, *European Sales and Marketing CMO Roundtable*, London, November 2006.
6. Chair, *Effective Innovation in the Food and Beverage Industry*, Amsterdam, December 2004.
7. Speaker, *MarketBridge Seminar*, London, March 2004.
8. Organizer, *Australian Marketing Institute: National Conference*, Sydney, September 1996.
9. Chairman, *Marketing and Strategic Planning in Education*, Sydney, April 1996.

10. Panelist, *The Learning Organization*, Sydney, July 1995.
11. State Councilor and Honorary Treasurer, Australian Marketing Institute, 1994-1996.
12. Member, Management and Administration Group, *Australian Institute of Management*, 1994-1995.
13. Rapporteur, *Best Practice in Management Development Conference*, Melbourne, May 1994.
14. Chairman, *Australia-Japan Business Outlook Conference*, Sydney, March 1993.
15. Panelist, *National Trade and Investment Outlook Conference*, Melbourne, June 1993.

Grants Received

1. Background research for a book on solution innovation (with Christen), INSEAD, 2013, €15,800.
2. The effects of virtual agents on customer usage of commercial websites (with Diesbach and Uusitalo), 2010, INSEAD, €25,625.
3. Research for examples to go into the book *The Innovation Manual*, INSEAD, 2008, €3,500.
4. Learning mechanisms and the development of dynamic capabilities within firms (with Soo, Cordery and *Alcoa Australia*), *Australian Research Council*, 2006, \$128,583.
5. Selection, Development and Production of a Platform for Multi-Agent Modeling Research (with Marks), INSEAD, 2005, €10,000.
6. Knowledge integration at the micro-level within the firm (with Soo), *INSEAD*, 2002, €2,803.
7. Determinants of e-business success: the impact of remedial efficiency and switching feasibility on the evolution of institutional structures (with Devinney), *INSEAD*, 2000, €23,700.
8. Benchmarking learned responses and rules of thumb in mature oligopolies (with Marks), *Australian Research Council*, 1998, \$17,537.
9. The organisational imperative and the performance of multinational firms (with Devinney), *Centre for Corporate Change*, 1998, \$30,000.
10. Implications of electronic commerce for organisational strategies and structures, *Centre for Corporate Change*, 1998, \$20,000.
11. Benchmarking learned responses and rules of thumb in mature oligopolies (with Marks), *University of New South Wales*, 1997, \$9,000.
12. Distribution of real and perceived publication rates in top marketing journals (with Dowling and Roberts), *University of New South Wales*, 1997, \$6,350.

13. A new model of global marketing, *Australian Research Council and the University of New South Wales*, 1997, \$13,000.
14. Conduct of a joint Indo-Australian symposium on enterprise development, *Australian International Education Foundation*, 1996, \$94,400.
15. Strategic behavior and learning: estimation of pricing and promotion in iterated oligopolies (with Marks), *Australian Research Council*, 1995-1997, \$75,000.
16. Organisational values and management development (with Kabanoff), *University of New South Wales*, 1995, \$6,000.
17. International patenting behavior and diffusion of a technological new product (with Devinney), *Australian Research Council*, 1994, \$11,000.
18. Influencing the diffusion of innovations within and between firms (with Roberts and Morrison), *Australian Research Council*, 1993-1994, \$20,900.
19. Strategic behavior and the evolution of economic institutions (with Marks), *Australian Research Council*, 1992-1993, \$20,200.
20. Development of an international marketing executive qualification, *New South Wales Education and Training Foundation*, 1992, \$86,393.
21. Provision of library materials on Pacific Rim markets, *University of New South Wales*, 1991-1992, \$10,000.
22. Identifying successful export marketing strategies for products based on scientific research, *Australian Research Council*, 1990-1992, \$165,025.
23. National Teaching Company Scheme (jointly with Burns Philp Pty. Ltd.), *Department of Industry, Technology and Commerce*, 1990-1993, \$50,000.
24. The role of communication networks in the diffusion of industrial innovations (with Roberts), *Australian Research Council*, 1991-1992, \$19,036.
25. The role of communication networks in the diffusion of industrial innovations (with JRoberts and Morrison), *University of New South Wales*, 1990, \$5,000.
26. Benchmark study of Australian management competencies and development needs, *National Board of Employment, Education and Training*, 1989-1991, \$61,000.
27. Forecasting the adoption of new high technology products (with Morrison and Roberts), *Australian Telecommunications and Electronics Research Board*, 1989, \$20,000.
28. National Teaching Company Scheme (jointly with MicroMedical Industries Pty. Ltd.), *Department of Industry, Technology and Commerce*, 1987-1988, \$30,000.
29. Establishment of the Centre for Export Marketing, *University of New South Wales*, 1987-1989, \$88,000.
30. Longitudinal study of market dynamics (with Dowling), *Australian Research Council*, 1979, \$9,000.

Education

1. Doctor of Philosophy, Marketing, University of Bradford, United Kingdom, 1974.
Thesis title: *a simple mathematical theory of innovative behavior*.
2. Master of Science, Management Studies, University of Bradford, United Kingdom, 1970. Thesis title: *the overlap problem in the UK TV market*.
3. Bachelor of Science, Pure Physics, University of Salford, United Kingdom, 1968
(awarded with *First Class Honors* and *Vice-Chancellor's Prize*).